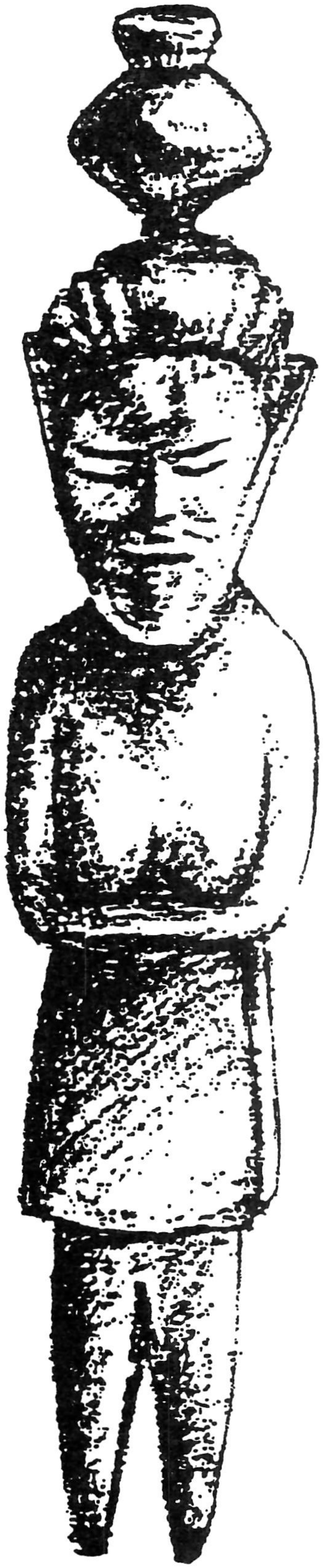
MOTHER TONGUE

JOURNAL OF THE ASSOCIATION FOR THE STUDY OF LANGUAGE IN PREHISTORY



Issue VIII, 2003

Special Issue: Proceedings of the Linguistic Databases and Linguistic Taxonomy Workshop
Santa Fe Institute, January 6-10, 2003

Edited by John D. Bengtson and George Starostin

CONTENTS

Introduction to *Mother Tongue* VIII

Main Articles

1 Inclusion Versus Exclusion: The Problem of Negative Evidence: *Paul Whitehouse*

9 Lexica Nostratica Addenda: *Václav Blažek*

21 Notes on Basque Comparative Phonology: *John D. Bengtson*

39 Salishan and North Caucasian: *Vitaly Shevoroshkin*

63 Toward Reconstructing Proto-South Khoisan (PSAK): *Christopher Ehret*

81 A Lexicostatistical Approach towards Reconstructing Proto-Khoisan: *George Starostin*

Informational Reports

127 Evolution of Human Language Project: Santa Fe Institute: *Sergei A. Starostin*

129 Linguistic Databases and Linguistic Taxonomy Workshop: *John D. Bengtson*

ASSOCIATION FOR THE STUDY OF LANGUAGE IN PREHISTORY
IN COLLABORATION WITH
EVOLUTION OF HUMAN LANGUAGE PROJECT (Santa Fe Institute)

Mother Tongue Issue VIII, 2003

Journal of the Association for the Study of Language in Prehistory
http://www.people.fas.harvard.edu/~witzel/aslip.html

OFFICERS OF ASLIP

As of Annual Meeting, April 19, 2003. Address appropriate correspondence to each.

President: Michael Witzel witzel@fas.harvard.edu
Dept. of Sanskrit and Indian Studies
Harvard University
1 Bow Street
Cambridge, MA 02138 U.S.A.

Vice-President: John D. Bengtson jdbengt@softhome.net
156 15th Avenue NE
Minneapolis, MN 55413 U.S.A.

Secretary-Treasurer: Peter Norquest norquesp@U.Arizona.EDU
1632 Santa Rita Avenue
Tucson, AZ 85719 U.S.A.

NB: during the absence of P. Norquest **in 2003/4**, please **send all fees**/contributions **to H. Fleming** (*)

***Mother Tongue* Editors:**
John D. Bengtson jdbengt@softhome.net
George Starostin gstarst@rinet.ru

BOARD OF DIRECTORS

Anne Windsor Beaman (Brookline, MA)
Ronald Christensen (Lincoln, MA)
Murray Denofsky (Somerville, MA)
Frederick Gamst (U. of Massachusetts)
Phillip Lieberman (Brown University)
Allan R. Bomhard (Charleston, SC)
Gyula Décsy (U. of Indiana)
Harold C. Fleming (Boston University)
John Robert Gardner (Marblehead, MA)
Michael Puett (Harvard University)
Jan Vansina (Madison, WI)

COUNCIL OF FELLOWS

Raimo Anttila (University of California, L A)
Luigi Luca Cavalli-Sforza (Stanford University)
Ben Ohiomamhe Elugbe (University of Ibadan)
Vyacheslav V. Ivanov (Russian Acad Of Sciences)
Winfred P. Lehmann (University of Texas)
Colin Renfrew (Cambridge University)
Vitaly Shevoroshkin (University of Michigan)
Ofer Bar-Yosef (Peabody Museum, Harvard)
Aharon Dolgopolsky (University of Haifa)
Dell Hymes (University of Virginia)
Sydney M. Lamb (Rice University)
Daniel F McCall (Boston University)
Merritt Ruhlen (Stanford University)
Sergei A Starostin (Russian State University)

© 2003 by the Association for the Study of Language in Prehistory
ISSN: 1087-0326 *Mother Tongue* **(Journal)**

* Harold C. Fleming, ASLIP, 16 Butman Avenue, Gloucester, MA 01930-1006, USA

Introduction to Mother Tongue VIII: Special Issue

This issue of Mother Tongue is a result of a Workshop held at the Santa Fe Institute in January, 2003. See the relevant reports for more details on the Institute and the Workshop.

The first division of the issue consists of six articles. In the first article Paul Whitehouse addresses an important theoretical issue. If it is taken for granted in the typical work on linguistic taxonomy that evidence is given for the *inclusion* of a language (or languages) in a family, what about all the evidence that other languages are *not* a part of the family?

In the second article Václav Blažek contributes yet another installment to the ever-growing corpus of evidence supporting the Nostratic macro-family hypothesis. It can be seen from Blažek's material that he prefers to hypothesize a "macro-Nostratic" that includes Afro-Asiatic, Dravidian, and Kartvelian, as well as the families included in Greenberg's "Eurasiatic." (For the latter see especially *Mother Tongue*, vols. VI and VII.)

The next two articles deal with another macro-family hypothesis, "Dene-Caucasian" or "Sino-Caucasian." Continuing a discussion begun in the first issue of *Mother Tongue* (Journal), John Bengtson presents lexical and phonological evidence for the inclusion of Basque in Dene-Caucasian. Vitaly Shevoroshkin offers striking lexical evidence connecting the Salishan (and Wakashan) languages of North America with the (North) Caucasian languages. This evidence could lead to the revision of the Dene-Caucasian hypothesis as well as the Amerind hypothesis; or at least shed significant light on migrations between Eurasia and the Americas.

The final two articles deal with a proposed macro-family that could prove to be even more ancient than any of the macro-families mentioned above. The Khoisan (macro)-family – although not yet universally accepted as a real family – is thought by some to be one half of a very old moiety: the split of Proto-Human into Khoisan and "non-Khoisan" (all the rest of the world's languages). Without making any sweeping pronouncements on this hypothesis, Christopher Ehret and George Starostin offer two approaches to the historical phonology of the Khoisan languages.

The final division of the volume consists of informational reports about the Workshop and Santa Fe Institute's program on the Evolution of Human Language.

Finally we express our deep gratitude to the Santa Fe Institute, and especially to Murray Gell-Mann and the MacArthur Foundation for making the Workshop possible.

Editors of this issue:
John D. Bengtson
George Starostin

Inclusion Versus Exclusion:
The Problem of Negative Evidence

Paul Whitehouse
Evolution of Human Language
Santa Fe Institute

It is standard practice for linguists, when they have identified the relationships between various languages and language families, to present evidence for such relationships in the form of collections of shared correspondences of sound and meaning. These are meant to demonstrate that the languages have a common origin; that they are all versions of a single common ancestral language.

By implication, this requires that all *other* languages are *not* versions of this ancestral language. But in our eagerness to look at the inclusive aspect of the relationships we lose sight of the exclusive — yet more than anything else it is the lack of evidence for the affiliation to the languages excluded that makes the proposal truly meaningful. This lack of evidence is never demonstrated, however, and so, in effect, the most important evidence for each proposal is routinely omitted. In other words, the best evidence is left out — and this is industry standard practice.

So, what does that mean for our understanding of genetic relationships? Firstly it reminds us of the relative nature of classification, and secondly it obliges us to think more carefully about what the writer is not saying. For instance, how far can we trust that other linguists have looked at all the alternatives and found nothing superior to the particular case they are presenting? Have they really looked everywhere? Were the alternatives really inferior and, if so, why? How do we know that the examples put forward (especially single attestations) are retentions from the proto-language(s) concerned, when the information *not* given may suggest a semantic history that conflicts with the one implied by the examples that *are* given?

The short answer is, of course, that we don't know. Unless we are as intimately acquainted with the primary data as the author is, we have to take all these things on trust. In Indo-European studies, for instance, where most of the participants are familiar with most of the subject matter, these things can, to some extent, be taken for granted. In less well-studied areas, however, or where we make comparisons between families we are familiar with and those we are not, the temptation to take on trust what our colleagues say is very strong. And there are strong inducements for us to succumb. We do not have time to become intimately acquainted with all the world's taxonomic questions, and if all taxonomists were to give each other the benefit of the doubt, the benefits of this in terms of convenience and a quiet life would make the temptation so great as to be almost irresistible.

But of course such an "I'll-agree-with-you-if-you-agree-with-me" approach is a scientific dead end, and, because it only takes one awkward voice to spoil it for everyone else by stating the obvious, we are left with no choice but to address the issue before someone else addresses it for us.

So, what does this mean in practice? Firstly, we need to be much more sceptical about what we are not being told; to read between the lines more. Secondly we have to be far more honest, both with ourselves and with others, about the merits of the alternatives to our own views. This is obvious, but rarely put into practice. Why? One obvious reason is shortage of space. However much we may wish always to present complete data, devoting the same space to alternative theories as to our own, even if those alternatives are simply "null hypotheses" whose supporting evidence consists only of scatterings of dubious comparisons, there is not a publisher on earth willing or able to devote sufficient paper to fit it all on to. I attempted to apply this approach to an article I wrote on the Nihali and Kusunda languages (MT III), giving every possible link I could find (however weak or contradictory), and it ran to 34 pages. The equivalent treatment of every language family in Australia, for instance, would run to hundreds if not thousands of pages And yet, how else are we to know that the case we are being offered is the best available?

As it happens, word processors and Internet transfer mean that we could actually disseminate documents of such scale amongst ourselves, but for anything beyond such *samizdat* publications, the lack of space available to us remains prohibitive.

The other constraint is simple human nature. We are all human, and human nature generally moves us to exaggerate the strengths of our arguments while minimising their weaknesses. We leave it to others to stress the weaknesses—but leave out the evidence that might make those others' task easier. I say "we" because I am as guilty of this as anyone.

But this is not the only way in which psychology impinges upon taxonomic linguistics. There are fundamental characteristics that make psychology an integral component of linguistic study.

Taxonomy, Psychology and Cognition

Linguistic change is an arbitrary and unpredictable process generating an endless sequence of unique phenomena, with no underlying physical basis from which we can formulate universal laws that might allow us to "deduce" linguistic prehistory theoretically. In the absence of absolute proof, all questions ultimately demand subjective judgements as to the best interpretation of the data, and all we have to rely on are educated judgements as to what is the most likely prehistorical sequence. Although we may seek to abide by common criteria, ultimately these judgements have to be subjective. In practice this means that debates are settled not by proof but by consent; not "what is the truth?" but "who agrees?"

So, why do linguists agree or not agree with particular genetic proposals? They agree on the basis of what they know. So, how do they *know* what they "know"? That is a matter of cognition, and that in turn depends as much upon the psychology of the observer as on the nature of the phenomenon under observation.

Language involves correspondences of sound and meaning. Thus the evidence we look for is auditory; we *hear* the evidence for common origin. However, the logistics of comparing great numbers of languages mean that the auditory evidence is stored using the written word and it is recalled visually; thus primarily we *see* the evidence of common origin. This is the essence of the multilateral approach to the initial stages of classification, where data from large numbers of languages are viewed all together. This is often dismissed as classification "by inspection," but there is no better place to *start* — and one could argue that everything else we go on and do to "prove" the relationships identified by "inspection" is little more than a *post hoc* rationalisation of the evidence of our own eyes.

Perception is subjective, and this has basic ramifications. We can only perceive for ourselves; we cannot perceive for others, and others cannot perceive for us. By the same token we can only know for ourselves; we cannot know something for someone else, any more than they can know something for us. Attempting to share our perceptions may even be counter-productive as long as we may be willing to accept a proposal from Professor A that we would not accept from Dr B, for no other reason than that B rubs us up the wrong way while A always lets everyone go ahead of him in the queue for coffee.

Finally, as well as seeing things differently, each of us is differently disposed to accept the evidence of our own eyes. For every linguist whose glass is half full, there is another whose glass is half empty. Little wonder then that linguists seem unable to agree on anything, except perhaps to agree that there is nothing out there that they can all agree on having seen. That is not to say that paleolinguistics is like some kind of Rorschach test in which "anything goes", but it *does* fix the point of decision firmly in the eye of the beholder. The best we can do is question our own perceptions more, and hope this brings us into greater agreement. And if not, let us be a little more understanding of what may sometimes seem like the hallucinations – or the blindness – of others.

Coastal South Papuan, A Test Case

At this point in the paper presented at the Workshop I sought to explore some of these issues via a test case I called Coastal South Papuan. This was a proposal that the Eleman, Kolopom and South Bird's

Head (SBH) families of southern New Guinea constitute a genetic node (CSPap). These widely separated families are three of the twenty-one co-ordinate branches of Stephen Wurm's Trans-New-Guinea phylum, whereas Joseph Greenberg, in his landmark 1971 article, placed South Bird's Head in his West New Guinea phylum, Kolopom in his South New Guinea, and Eleman in his East New Guinea.

I approached this question by examining in detail Greenberg's evidence with respect to these three families, plus a briefer look at the basis for the Wurm classification, and offering evidence for my own Coastal South Papuan alternative in the form of sets of words linking its three branches. Finally, I also gave sets of etymologies for each of the three branches, to allow for a better understanding of their respective phonologies.

When it came to publishing the paper in *Mother Tongue*, however, this section immediately fell foul of technical problems, specifically the incompatibility of my computer set up (Apple OS) and John Bengtson's (Microsoft DOS). This was further complicated by the extensive use of specialised fonts, which requires both parties to have compatible software. This problem will persist as long as linguistics involves the use of specialised fonts, *i.e.* forever. I decided that the simplest way round this problem was to omit the CSPap section entirely. But, even if my system and John's had been perfectly compatible, I would still have had to omit three of the four appendices in order to comply with the 50 page limit for papers.

My real reason, though, for accepting this solution so readily (and if this paper is about anything, it is about looking at how the truth may be variously concealed, distorted and ignored), was a loss of confidence in my own data.

My fall-back position when I began to put the material together was that the paper could just as easily be about why CSPap did not work as about why it did, but now that the first looks increasingly more likely than the second I find myself increasingly less indifferent as to which it is to be. I am now happy just to discuss the pluses and minuses without the supporting examples. This is a poor substitute for the real thing, but I have good reason for prefering to describe the inadequacies of my paper rather than letting you witness the full horror for yourselves.

The Whitehouse Method (1): Conception

First of all, how was it that I came up with such an unorthodox proposal? It was immediately obvious when I began to compare them that the Greenberg and Wurm classifications were so completely different as to be mutually exclusive. I therefore thought it best to assume that both schemes were wrong and set out to partition the languages for myself. This seemed like good science; if my work confirmed one or other of these schemes in a "blind test," good. If they suggested a third scheme, equally good.

Using the Wurm classification as a guide to the lowest level relationships I entered data for each genetic unit on a series of crib sheets, a column for each unit, 35 lexical items to each crib sheet (it was necessary to complete a second crib sheet). These are all parts of the core vocabulary ('blood', 'hair', 'louse', 'stone', etc.). The data entered were forms representative of the genetic unit concerned. Where there was a single obvious shared form this alone was entered (with sufficient alternatives to suggest its original form); where there was not, all possible alternatives were set down. Thus the data I used were already "sorted," to reduce the possibility of using forms that were not representative of the ancestral language concerned. To ensure that no languages had been mis-allocated by Wurm I took a very conservative view as to what belonged with what, and so only reduced the mass of languages down to 200 or so obviously separate units.

These data were then entered on a second set of crib sheets, one for each lexical item, according to their consonant pattern; this crude arrangement allowed me to see how forms were shared across the whole range of genetic units. One strength of this approach is that at the outset all units were compared equally, regardless of geography or reputation.

I expected this exercise to confirm either Wurm or Greenberg (with Greenberg the favourite), but it did neither. Indeed, the absence of *any* clear fault lines within the New Guinea languages was striking, and it was this that necessitated the compilation of the second set of crib sheets (doubling the 35 or so lexical items to 70-odd). Even then the confusion was only partially alleviated. However, neither Greenberg nor Wurm were confirmed by the second set of data, any more than they were by the first. My preliminary conclusion was that the subgrouping of Indo-Pacific would turn out to be much more complex than the (apparently) zonal scheme proposed by Greenberg, with subgroups arranged in ribbon-like patterns, sometimes even linking groups found on opposite sides of New Guinea. The distribution of South Bird's Head, Kolopom and Eleman scattered along the south coast is typical of such patterns.

One other obvious deficiency of this process is that it does not take into account possible links to non-New Guinea families. However, my object in the first instance was not to subgroup Indo-Pacific, but merely to boil it down into few enough chunks to be used in global comparisons that may or may not confirm (and sub-group) Indo-Pacific as part of a global classification.

This methodology goes a long way towards explaining these unorthodox findings, and the reason lies as much in what I did not do as in what I did. I did not look at either Greenberg's or Wurm's or anyone else's work while I was trying to identify patterns within the mass of data. I also avoided looking at the map until I was confident that I had begun to make up my own mind. Indeed, it may be that a proposal such as Coastal South Papuan could not have been conceived by anyone who *had* looked at the map first. My global comparisons have followed this same principle, of comparing everything against everything else regardless of geography.

Because the human language family is unique and finite, such thoroughness seems obvious. After all, with only one set of languages to classify, the job only has to be done once. And, once everything has been compared to everything else, it will no longer be possible for others to object that evidence linking particular groups of languages could be matched by equivalent evidence linking various of these language groups with different language groups, since we can then say with confidence that we have looked for such equivalent evidence and not found it.

Obviously such all-inclusive comparisons may throw up a range of chance resemblances, but this is no bad thing. People talk readily of "junk hits," but no one ever takes the trouble to lay all the "junk" out for inspection, and so no one ever comes to understand what "junk" really looks like. But for us not to do it would be like an archeologist denying the necessity of dirtying his boots down some hole because there's nothing down there but dirt.

The Whitehouse Method (2): Presentation

Here also, what I did not do is as significant as what I did. The method outlined above is clearly designed for the initial stages of classification, the generation of hypotheses. As such it has been a great success, but that does not necessarily mean that the hypotheses themselves are equally successful, nor that the method described is as good for validating hypotheses as for generating them.

The first thing that would strike the reader (if s/he had a copy of the CSPap etymologies to read) is that they are preliminary in nature. In some cases they are no more than collections of words whose CVC patterns match crudely according to shared points of articulation. The purpose of this was to allow me find recurrent sound correspondences. Beyond a couple of blindingly obvious cases, I made no attempt either to identify those correspondences or to edit (and discard) accordingly. Thus the etymologies include a great many putative cognates that are unlikely to be valid, if not inherently invalid.

My argument has always been that it is the rubbish that makes the gems shine out, but in retrospect this is only once they have been cleaned and polished. Otherwise, the response of the reader is more likely to be, "Gee, look at all that rubbish!"

In addition to the unrefined primary data, I also included those reconstructions from Austronesian and its subgroups that seemed to offer an alternative, plus the corresponding words in Motu, which is widely used in southern New Guinea as a *lingua franca*. The purpose of this was to address the possibility of shared forms being the result of borrowing from some common source. The underlying statement was, "This is the best alternative I could find. If no alternative is offered, it is because I could not find one. If you think these are unconvincing, you should see the ones I discarded." In practice, however, this is more likely to confuse the reader, who never quite knows whether the author is suggesting that these are supposed to be cognate or not.

My intention was to offer the reader sufficient options from which to choose and enough information with which to make up his own mind, rather than simply telling him what to think. Perhaps this was a mistake. As Tim Usher pointed out to me, having examined the case for CSPap in detail, the trouble with weak data is that people stop reading it very quickly.

I have already touched upon the question of sound correspondences, but in addition to expanding upon this there are some other issues to consider first.

Semantic Histories

Every word that appears as part of a wider set must, *ipso facto,* be a retention from the ancestral language of the subgroup to which it belongs. This sounds obvious, but it has implications that are often overlooked. Firstly, every time an etymology includes a single word from one of the component families, that word is an etymology in its own right and by implication it belongs in any collection of etymologies for that component family.

This is important because if we do not, we may later be tempted to use a different word from a different member of the component family in some other wider etymology. Logically these would both have to be retentions from the ancestral language, yet neither would appear in the collection of etymologies for the component family as typically presented. The danger here is that we may lose track of just how many different retentions we are postulating for a single lexical slot. This raises the question of how many synonyms are allowable for a proto-language, and also what is the relationship between these synonyms at different stages of language history.

Continuity is the key. Words put forward as retentions from the ancestral language must *always* have been there, along side whatever other words have also been put forward as retentions from the ancestral language, at *all* stages of its development. Each must justify its place along side all the others; we must have reasons for accepting all of the alternatives. If we are not able to construct satisfactory semantic histories to account for all of the words being put forward as retentions from the ancestral language, it must give rise to the suspicion that one or other of the alternatives is a chance resemblance and not a true cognate.

Insofar as the etymologies for Coastal South Papuan and its component families, which appeared as four separate appendices, were originally arranged (and examined) as a single collection, this problem was to some extent accounted for.

I also attempted to analyse the cases where Greenberg used single attestations in his evidence for Indo-Pacific and its subgroups, suggesting CSPap alternatives that might imply that the Greenberg examples might be borrowings. Insofar as this was valid, it was undermined by the failure on my part to give the data that would demonstrate different relationships, both for the component families of CSPap and for the families from which I sought to separate them. One reason is that I had not yet been able to complete the detailed work necessary, the other is that there would not have been room for it here.

This sort of problem bedevils Indo-Pacific classification, with many of the sub groups having more than one possible affiliation, though it may also be seen wherever else the industry practice of publishing "inclusive-only" evidence (as followed by Greenberg, for instance) means that data can only ever be presented out of context.

What it does do is highlight the difficulties inherent in presenting negative evidence. The problem may be insoluble as long as there is no opportunity to publish every possible link between every possible pair of families, but it may at least make linguists more inclined to bear in mind what has not been included.

Glottochronology and Lexicostatistics

It was largely on the basis of mathematical techniques that Wurm arrived at his classification of the non-Austronesian languages of the New Guinea area. This is mostly successful for delineating the lowest level relationships, but it also makes some relationships look closer than they appear when you look at the evidence in a more traditional way (the placing of Kayagar, Asmat-Kamoro etc). Unfortunately it is only at the time depths where these mathematical analyses break down that taxonomy really becomes interesting.

Wurm postulated 21 coordinate branches for his Trans-New Guinea phylum; to address this proposal consistently would require me to present and discuss the evidence linking SBH, Eleman and Kolopom to each of these other 18 families, but even if I had done the work, there would not be room to fit it all into a paper like this.

Typology and the Trans-New Guinea Phylum

So far as I understand it, the higher level classification(s) published by S. A. Wurm are in fact as much typological as lexicostatistical, even though neither of these are satisfactory guides to deep-level relationships. Yet it is on the basis of typological criteria that SBH is removed from the West Papuan groups variously put forward by Cowan and Greenberg, while the proposed affiliation of Eleman with the East New Guinea Highland Group seems to be based mostly on oral tradition (Franklin, in Wurm 1975, pp.859-862).

I cannot help thinking that when a linguist has recourse to typology and substratum influences it is a sure sign that they are too daunted by the complexities of the phenomena before them to do the work properly. Arguments of this kind merely confirm that typology generates bad taxonomy, and it is only when linguists get to grips with specific correspondences of sound and meaning (whether borrowed or inherited), and recognise that substratum influence ultimately involves specific exchanges between identifiable pairs of languages, that the true relationships will ever be identified.

Sound Correspondences

Sound laws are like cognate sets insofar as linguists habitually only present the evidence that fits the theory and leave out the rest, thus discouraging the investigation of alternative explanations. More than anything else they provide an excuse for readers not to look beyond the evidence presented, but here too they require us to ask how far we can trust another linguist to have done our looking for us. Another problem is, of course, that our ability to identify sound laws is limited by the amount of data available. The best sound laws come from the most obvious relationships; one might almost say that the better the sound laws, the less we need them. Certainly sound correspondences only *really* become useful beyond the point where it starts becoming *really* difficult to identify them.

The identification of sound correspondences even within the three component families under review here is made difficult by the limitations of the data available. For instance, it is often hard to know what are the Purari and Tate reflexes of quite common correspondences within Eleman simply because the Purari and Tate equivalents are not cognate. Sometimes, for the really important correspondences, there is only one match, and a single sound correspondence intrinsically can *not* be recurrent. Unfortunately this sort of bad luck is inherent in the comparison of languages (like Tate and Purari) that are relatively remote from their closest relatives. Sometimes, the arbitrary nature of language change means that languages with, say, 20% vocabulary in common can produce better sound correspondences that languages with 30% or 40% in common. We can only hope that we will one day be able to obtain word lists long enough for us to fill in the gaps.

This limitation affects all three families. For instance, my appendices included far too many cases where the sole attestation in SBH is Weriagar, or the sole attestation within Eleman is Orokolo, simply because these are the only languages in which my word lists include the lexical item concerned. We particularly need longer lists for Konda, Yahadian, Duriankere and Tate.

But also bear in mind that these are deficiencies you would not be aware of if I had not told you of them.

I must also admit that my respect for sound correspondences is much greater now that it was before I began looking for them, and the failure to address those correspondences that *were* identifiable resulted in the inclusion of many comparisons whose deficiencies made my overall case weaker than it might have been.

Health Warning

This part of the workshop paper addressed the possibility that because the comparative exercise from which this paper was drawn was still far from complete, I may have to modify these proposals at some stage — and so it has turned out. Tim Usher has produced a detailed commentary on my paper based on work that is more advanced than mine, which I have not yet steeled myself to examine in equal detail. All I would say at this stage is that if anything of Coastal South Papuan survives Tim's onslaught, it really will have to be taken seriously.

It also goes without saying that, however valuable not looking at other linguists' work may be at the early stages of hypothesis formulation, there comes a time when it can no longer be put off. That moment should, by rights, come before the publication of a proposal like Coastal South Papuan, not after it.

Conclusions

So where does this leave us in our search for the optimum balance between inclusion and exclusion? For the moment it looks as though the limitations of space, both on the page and in the short-term memory of the neurologically typical linguist, will continue to oblige is to maintain a narrow focus. It is questionable whether this situation will ever change. In the meantime, we will just have to try to remain alert to our own blind spots, both individually and as a discipline.

Acknowledgements

Tim Usher and I have freely exchanged the data that we have amassed in our respective investigations, and this has been to our mutual benefit. It should not need saying, but such pooling of information should be the norm rather than the exception. In addition to Tim's selfless willingness to share with me the mass of data he has accumulated down the years, I have also received many, many copies of wordlists from the Summer Institute of Linguistics office in Ukarumpa, Papua New Guinea. These unpublished data have filled in many of the worst gaps in the material available in print and I cannot speak highly enough in praise of the Office Manager there, Ruth Thurman, who has never been less than helpful and enthusiastic.

Here at least is one statement that I have no hesitation in repeating in print.

References

Greenberg, Joseph H., 1971, 'The Indo-Pacific Hypothesis,' in *Current Trends in Linguistics,* 8, ed. by Sebeok, Thomas A., The Hague, pp.807-870

Wurm, S. A. ed, 1975, *Papuan Languages and the New Guinea Linguistic Scene*, New Guinea Area Languages and Language Study, Vol 1., Pacific Linguistics C-38, Canberra

Lexica Nostratica Addenda

Václav Blažek
Institute of Linguistics, Faculty of Arts
Masaryk University
Brno, Czech Republic

The following list of Nostratic etymologies was compiled during the summer and autumn of 1989. After the fundamental change of *ancien régime* in my country in November 1989 I did not return to it for more than thirteen years. Stimulated by two conferences devoted to the problems of distant linguistic relationship organized by Sergei Starostin and Murray Gell-Mann at the Santa Fe Institute, New Mexico (Oct 2002, Jan 2003), I took advantage of hospitality of the Netherland Institute of Advanced Study in Wassenaar to choose those etymologies that were not published in my previous articles (e.g. 1992, 2002), with the exception of significant supplements.

Concerning the basic correspondences in consonantism, I follow Illič-Svityč (1971), naturally taking into account progress in reconstruction of particular proto-languages, especially Afroasiatic and its branches. In reconstruction of the Nostratic vocalism I am inspired by the ideas of Sergei Starostin, Anna Dybo and Oleg Mudrak, as applied to their reconstruction of proto-Altaic. The title reflects a continuation of my previous articles devoted to the additional material in the Nostratic lexicon (Blažek 1989, 1990).

1a. ***'in(ʌ)ši** "man / people // woman"
AA: Sem ***'inaš-** "man": Hbr **'ănōš** "man, mankind", Aram (Syr) **'nāš-ā**, Arab **'ins** "man", ESA **'ns**, Jibbali **ènsí** "human, not jinn", Tigray **enās** "homme" | Om (N) Mao **nuuše** "husband" | Berb: Tuareg of Ahaggar **a-ynəs** "young man" (from data presented by A. Militarev at 5th International Hamito-Semitic Congress, Vienna 1987)
U ***inše** "Mensch" (SKES 102; UEW 627-28: without Samoyed data);
Note: See Koskinen 1980, 130-31: Sem + U;
cf. 1b. Sem ***nisw-** "woman": Aram **n^{e}šayy-ā**, Syr **neše**, Arab **nisw-at** | Cush: (E) Jiddu **na'as**, Boni **nasa** "woman" (Dolgopolskij 1973, 289) | Ch: (C) Tera **nušu** id. (Newman), Pidlimdi **nušu**, Glavda **nùsà** id. (Kraft)
U: FV ***nisз** "Weib, Frau" (UEW 708).
Are these forms related or different?

2. ***^{c}eŕʌ** "rainy cloud"
AA: Cush: (C) Awngi **ərí** "rain" (Hetzron) | (E) Afar **^{c}ir** ~ **^{c}er** "smoke" (Parker & Hayward), Somali **^{c}iiro** "fog, cloud" (Reinisch), Arbore **'irí** (Hayward), Elmolo **íri** (Heine), Dasanech **'ir** "rain" (Sasse), Harso-Dobase **iráwwo**, Gollango-Gawwada **irráwo** "rain" (AMS), Tsamay **'a̟rro** id. (SLLE) | (Om: (N) Wolayta, Zala, Gofa, etc. **ira** (Cerulli), Kullo **iira**, Dache **iyra**, Zayse, Zergulla, Gidicho **iira**, Yemsa **iiro**, Dizi **iiru**, Shako, Nao **iru**, etc. "rain" (Bender) - see Dolgopolskij 1973, 143 | Ch: (W) Ron: Daffo-Butura **ṛí'** "cloud", Sha **'aré** "cloud, sky" (Jungraithmayr); (C) WMargi **ʾyar**, Kilba **'e̱r** "smoke" (Kraft)
D ***eṛ̤-** "cloud" (DEDR 849).

3. ***bädü** ~ ***badi** "many"
AA· Cush: (E) Saho **bad-** "to be rich", Som **badi** "to be numerous" (Reinisch) | Om. (N) Kafa **bad** "to become rich" (it is not excluded that some of these forms have Ethio-Semitic origin, cf Tigre **bit** "to be rich"; see Dolgopolskij, *Jazyki Afriki*, M. 1966, 52); (S) Ubamer, Galila **bɛdi**, bako **bɛdimi** "many" (Fleming); Ehret (1980, 320) adds Alagwa (SCush) **bar-is-** "to grow (of plants)"
Kartv ***bed-** "fate, luck" (Fähnrich, *Georgica* 4, 1981, 92)
IE ***b^{h}od^{h}-**: Welsh **bodd** "goodwill, consent", Middle High German **bat** "help, use" (Mann 1984-87, 88)

A: Tk ***bāδū̄-**"to become big, high; to grow", cf. Turkish dial. **bed** "many" (VEWT 67; Sevortjan II, 288) | Mong **büdügün** "grob, dick" (KW 66) | Tg ***badï-** "still more / better" (TMS I, 63).

4. ***budi** "ashes"
AA ***bud-** "ashes" (SISAJ I, 49-50): Om ***bud-(in-)**: (S) Dime **bIndo**, Ari **bIndi**, Banna **bIddIni**, Karo **bidɨni**, Hamer **bidini** etc. "ashes" (Bender / Fleming); (N) Malo **budó**, Dokka **buda** (Fleming), Basketo **buda**, Wolayta **bidinta**, Kullo **bidIntsa**, Oyda **buudɔ**, Dorze **bu̲ddu**, She **ben**, Yemsa **bɛdnə** etc. id. (Bender) | Ch: (W) Hausa **hábḍìì**, Karekare **butaw**, Diri **bùtù**, Ngizim **bébéḍ**, etc.; (C) Yedina **budén**, Masa **búdu**, etc ; (E) Dangla **bùtù**, Migama **bíttí**, Birgit **bùtì**, etc. "ashes" (JgIb 4-5)
D ***poṭi** "dust; become pulverized" (DEDR 4481)
?A: WrMong **buǯar** "dirt; dirty" (KW 64; cf. TMS I, 103).

5. ***buḥrʌ** ~ ***burḥʌ** "lake / river / brook / marsh"
AA: Sem ***baḥr-** "sea, river" (Cohen 1970, 56)
IE ***b^{h}ōr-**: Sl ***bara** "brook, marsh, pool" (ESSJ 1, 153-55); cf. also Gmc ***brōka-** "brook; swamp", etc. (Kluge 1975, 103)
U: BF: Finnish **puro** "Bach", Karelian **purakko** id. (SKES 655)
D ***pur̤-** "river" (DEDR 4318)
A: Tg: Evenki **buruk** "whirlpool" (TMS I, 114).

6. ***buti** "louse"
AA: Cush: (C) ***bət-** "louse" (Appleyard)
D ***putir̠-kV** "louse, gnat, midge" (DEDR 4203)
A: Tk ***bijt** "louse" (VEWT 76).
Note: Dolgopolskij, *Voprosy jazykoznanja* 1964/2, 61: CCush + Tk.

7 ***ćilu** "all"
U· FP ***ćilз** "alles; ganz" (UEW 613)
A: WrMong **čulu**, Kalm **tsul^{u}** "ganz" (KW 433).

8. ***čajku** "back"
AA ***čVkVm-**: Sem ***t̠akam-**: Ug **t̠km** "Nacken mit schulter", Hbr **šəkäm** "shoulder, nape of the neck, back" (SED I, 251-52)
IE ***(s)teig^{w}-**: Arm **t'ēkn** "Schulter, Achsel, Arm", pl. **t'ikunk'** "Schultern, Rücken, Hinterteil, Seite"; OIr **tóeb**, **tóib** "Seite", Welsh **tu** id. (P 1018)
A: Tk ***čaikan-** "ell" (VEWT 96, 107; Dybo 1996, 169-72) || Tg ***čaxa(n)** "ell, arm-pit" (TMS II, 378, 380; Dybo 1996, 311).

9. ***dar^{c}a?** "ashes"
AA: Cush: (E) ***dar^{c}-** "ashes" (Sasse 1979, 16) | (S) Iraqw **dá^{c}árítô'**, pl. **da^{c}ara** id. (Whitley)
D ***tarampu** "ashes" (DEDR 3092).

10. ***garCi** "louse; flea"
AA: Cush: (E) ***'azgir-**: Som, Rendille **injir**, Jiddu **'ajjere**, Bayso **iggir**, Arbore **'ingir**, Elmolo **íŋgir**, Dasanech **izire**, Oromo **injiran**, Konso **íkkirteeta**, Dirayta **ikkirét**, etc. "louse" | Om: (S) Dime **garso** id (Bender), (N) Gofa **aŋgarc̣o** "sp. insect" (Fleming) | Ch: (C) Muzgu **ngə́rsa** "louse"; (E) Kwang **ngə̀rsà**, Mokilko **gérsè** id. (Mukarovsky 1987, 249)
Kartv ***g̥r̥c̣q̇il-** "flea / Pulex irritans" (Klimov 1964, 65).

11. ***gok'a** "top (of head)"

U: FU ***kokka** "etwas Hervonstehendes, Spitze" (UEW 171-72) | Yukaghir **kókä** "Kopf" (Bouda, *Ung. Jb.* 20, 1940, 76: FU + Yukaghir)
D ***kukk-** "head, extremity" (DEDR 1630)
?A ***góki(dV)** "high, peak": Mong ***gögde** (KW 137) | Tg ***gugda** "high" & Manchu **gukdu** "peak, hill" (TMS I, 166, 169) | MKor ***kòkái** | pJp ***kúkì** (AED).

12. ***gon(j)ʌ** "kill"
AA: Sem **√g-n-y/ ʼ**: Akk **genû** "stossen", Hbr-Aram **gnʼ** "concussit" | Eg **gn.t** "wound, slit" | Cush: (E) Som **gan-** "tirar con arco", Sidamo, Hadiyya **gan-** "to beat" (cf. Dolgopolskij, *Problemy afrikanskogo jazykoznanija*, Moskva: Nauka 1972, 211; Bomhard 1984, 244)
IE ***g^{wh}en-** "to kill" (P 491-93)
A: Mong **gönen** "die Töten" (KW 138).

13 ***gori** "long / far / high"
AA: Ch. (W) Ngamo **gara**, Bole **garan** "long"; (C) Bachama **guragura**, Nzangi **gɨrgɨr** id. (Mukarovsky 1987, 248)
Kartv ***gr̥ʒel-** "long" (Klimov 1964, 65)
?IE: WGmc ***grauta-** "big" (Kluge 1975, 272)
U: FV ***korkз** "hoch" (UEW 672)
A ***gure**: Tk ***gür** "thick, broad" (VEWT 310-11) | Mong ***gür** "wide, broad" | Tg ***gora** "far, long" (TMS I, 161-62) | MKor ***kūrk-** "thick" (AED).

14. ***gubi** "to burn, fry, cook"
AA: Sem: Akk **gubbubu** "to roast" (cf. Dolgopolsky 1983, 132: Akk+ECush) | Cush: (E) ***gub-** "to burn" (Sasse 1979, 17)
Kartv ***gab-/*gb-** "to cook" (Klimov 1964, 58)
?IE ***g^{wh}ob$^{(h)}$-**: Old Lithuanian **gabija & gubija** "fire" (in a noble speech), **Gabija** "fire-goddess" (Lasicki, 1615), and maybe Latin **faber** "smith", Gaulish (Alise-Sainte-Reine) dat. pl. **gobedbi**, Welsh **gof**, OIr **goba(e)** "smith"
U: FP ***küpe(-ńз)** "Funke" (UEW 665)
A ***gǖbe**: Tk ***gübeč** "frying pan" | Tg ***gǖw-**: Evenki **gī-** "to fume, smoke", **gīvun** "furnace, stove", Manchu **guva-x́an** id. (TMS I, 147, 165) | MKor ***kūp-tá** "to bake, fry" | pJp ***kùjú-r-** "to smoke, fume" (AED).

15. ***gudi** "guts / bowels / belly / inside"
AA: Cush (E) ***gidd-/ *gudd-** "inside, centre, middle" (Sasse 1979, 16, 18; Dolgopolskij 1973, 239)
U: FU ***kütз** "Mitte, Zwischenraum" (UEW 163)
D ***kuṭ-** "bowels, intestines" (DEDR 1652)
A ***giodV**: WrMong **gede-sün** | Tg ***gudige** "belly, stomach" (TMS I, 167; AED).

16. ***guŋija ~ *gijaŋu** "day- or night-light"
U ***kuŋe** "Mond; Monat" (UEW 211-12) | Yukaghir S **kiniʒe**, N **kinʒe id.**
A ***gi̟òjńu** "dawn, daylight": Tk ***güń(al')** "sun; day" | Mong ***gege-γen** "dawn, daylight" (KW 132) | Tg ***giańam** "dawn" (TMS I, 145) | MKor ***kúi** "dawn" | pJp ***kai** (AED)
Esk ***qaŋu**: Greenland **qaummat** "moon", EEsk **qau-** "to be daylight" (Bergsland, *JSFOu* 61, 1959, 13).

17. ***guri** "mountain, hill / rock, stone"
AA: Sem: ESA **grb** "campus montanus consitus; saxum rude, lapis ruber" (Conti Rossini) | Cush (C) Qwara **gara** "rock" (Reinisch); (E) Som **guuro** "peak", Oromo of Tulama **gaaraa** "mountain", Sidamo **gaaroo** id , Burji **gori** id.; (S) Iraqw **gâra**, pl. **gaḍu** "forest" (Whitley) | Om: (N) Shinasha **guuroo**, Anfillo **gura**

"mountain" (Dolgopolskij 1973, 61) | Ch: (W) Mburku **gírí** "stone, rock"; () Munjuk **gòrhòyô** "stone"; Migama **gûrbú** id., Dangla **gìrpì** "rock" (Mukarovsky 1987, 357)
Kartv ***gora-** "hill" (Klimov 1964, 64)
A: Mong **güri**, **gürü** "grosser Stein" (KW 139) || Jp **iši-koro** "stone, pebbles" : **iši** "stone", cf. Mong **gürü-čilaγu** "pebble" (Miller, R.A., *The Japanese Language*, Chicago 1980, 71-73).

Note: U/FP ***kurз** "Hügel, Anhöhe, Landrücken" (UEW 677) and D ***kur̲-** "hill country" (DEDR 1844) can belong here too, or to IE ***g^{w}erH-** "mountain" & AA: Sem: ESA **kwr** "hill", ARAB of Hadramawt **kawr** "mountain" | Cush: (N) Beja **kaar** "hill"; (E) ***kor-/*kur-** "mountain, highland", e.g. Elmolo **kóóran** "mountain" | Ch: (C) Nzangi **kùròmo** "mountain", indicating Nostr ***kurʌ**.

18. ***gulʌ** "fight / destroy"
AA: Sem: Arab √ **ġ-w-l** "to kill, perish", etc., ESA **ġll** "be filled with hatred" (Biella 1982, 395) | Cush: (E) ***^{c}ol-** "war" (Sasse 1979, 21; Dolgopolskij 1973, 162)
IE ***Hwol-**: Hit **hullai-** "to smite, destroy"; Greek ὄλλυμι "I destroy"; Lat **ob-oleō** id. (P 777)
D ***ul-** "to ruin, perish, be wasted" (DEDR 671).
Note: Bomhard 1984, 58: AA+IE.

19. ***hila** "moon / light"
AA: Sem √**h-l-l**: Ug **hll** "Neumondsichel", Hbr **helēl** "morning-star", Arab **hilāl**, Geez **helāl** "Neumond" (Aistleitner 1963, 89; Cohen 1970f, 415-17) | Berb: Tuareg of Taitoq, Ahaggar **tallit**, pl. **tilil** "new moon" | ?Cush: (E) Saho-Afar **alsa** "moon, month" (Reinisch)
?IE: Cuneiform Luwian **halli-** "day", Hieroglyphic Luwian **hali-** "day, time"
D ***el** "light, sun" (DEDR 829)
A: Tg ***ila-** "(to) light" (TMS I, 303-04).

20. ***HanHa(tʌ)** "sp. duck"
IE ***H$_2$enH$_2$t-** "duck" (P 41-42)
U ***aŋa** id. (UEW 13)
A ***ăŋatV** "sp. duck": Tk ***ăŋ(k)ɪ̄t** (VEWT 21) | Mong ***aŋgir** | Tg ***(x)andi** (TMS I, 43) | pJp ***anti** (Starostin 1991, 71; AED).

21. ***Hari** "pond, tank"
AA: Sem: Akk **ḫarru** "watercourse" | Cush: (E) ***har-** "pond, brook, river" (Sasse 1982, 91)
IE ***H$_2$eri-**: Armenian **air** "spelonca, grotta, caverna", Hittite **hari-** "valley" (B. Čop, *Indouralica I*, Ljubljana 1974, 32: IE + FU)
U: FU ***arз** "wasserige, sumpfige Stelle" (UEW 17)
D ***ēri** "tank, lake" (DEDR 901).

22. ***Hulʌ** "all"
IE ***ol-** "each, all" (P 24; M 872): Welsh, Breton **oll** "great, ample", Old Irish **oll** "great, ample" (***olno-**), comp. **(h)uilliu** "amplius" (***ol-yōs**), Gaulish *ollon* nom./acc. sg. ntr. "large, grown up", Goth **alls** "all" (***olno-**), **ala-mans** "all men" (***olo-**); cf. Toch B **olya** "more" (comparative ***ol-yōs**)
U ***ulз-(jз)** "viel, gross; sehr" (UEW 543-44); cf. Nenets **ŋūl'ī'** "sehr, überaus; ganz"
A ***ulu/o** "big; many; good": Tk ***ulug** "big" (VEWT 513, 520) | Mong ***olon** "many" (KW 285) | Tg ***ule-** "good" (TMS II, 260-61) | MKor ***ōr-** (AED).

23. ***ḥoma** "bad"
AA: Cush: (E) ***ḥam-/*ḥum-** "bad" (Sasse 1979, 38) | (S) Burunge **ḥam-**, Alagwa **hamu** "hardship, distress" (Ehret 1980, 379, 334)

IE ***aumo-** or ***oumo-**: OIc **aumr** "arm, elend", Toch B **aume** "misère" (Van Windekens 1976, 154, 334-35, 539)
U: FU ***oma** "weak, old; evil ghost" (UEW 337)
D ***ūm-** "dumb" (DEDR 746)
?A: Mong: Kalm **omūn** "weich" (KW 286)
Nivkh **um-la**, **-ra** "böse, zornig" (Bouda, *Anthropos* 55, 1960, 109).

24. ***jer(j)u**? "river"
AA: Berb: Zenaga **tiert** "river" (Basset: **√y-r**) | Om: (N) Anfillo **yuro** "water" (Bender) | Ch: (C) Muzgu **ẹrẹ́** "fleuve" (Mouchet)
?IE: Balt ***yeurio-** "sea": Lith **jū́rės, jū́rios** pl., OLatv **jūri-**, Prussian acc. **iūrin**, pl. **iuriay** (P 81)
D ***yār̲u** "river" (D 5159)
A: Tk ***jïrmak** "Fluss, Strom" (VFWT 202: Mong **irmaG** "Kante, Rand, Ufer"), cf. East Turkic **iren** "water", Osman dial. **irim** "bay", **yran** "flowing", etc. (Sevortjàn I, 664-65).

25. ***kana** "feather / wing"
?AA: Sem ***kanap-** "wing" (EDE 130-31) | Cush: (C) ***kanf-** "wing, feather" (Dolgopolskij 1973, 65)
D ***kēnt-** "feather, hair" (DEDR 2002)
A. Tk ***kānat** "Feder" (VEWT 230) | Mong **kana** "grosse Feder" (KW 165)

26. ***käC'ʌ ~ *k‘äz/ʒʌ** "cold"
AA: Sem: Akk **kaṣū** "cold" (***ç̣/ḉ̣/ç̣̌**) | ?Cush: (C) Bilin **qaž-qaž**, Khamir **qaz-qaz**, Qwara **xexez** "cold" (Reinisch) | (E) ***qiiz-** "cold" (Sasse 1979, 49)
U: FP ***käćɜ** "kalt werden, sich erkälten" (UEW 648).
Note: Regressive or progressive assimilation according to glottalization?

27. ***k**$^{(‘)}$**oli** "long, far"
D ***kōlu** "long, large" (DEDR 2239)
A ***k**$^{(‘)}$**i̯olo**: Mong ***kolo** "far" (KW 182) | MKor ***kìr-** "long" (AED).

28. ***körʌ** "grass"
AA: Ch: (W) Hausa **kooree** "green", Tangale **kor̃ok** "grass"
D ***kīr-** "greens, vegetable, herb" (DEDR 1617)
A: Tk ***köräŋ** "Riedgras / carex flava" (VEWT 292).

29. ***kul/l'ʌ** "ashes"
IE ***geulo-** "coal": Irish **gúal** id., Gmc ***kula(n)-** id. (P 399), Toch B **śoliye** "ohniště" (***geulu-H**$_1$**en-**) (Adams 1999, 635)
U: FU ***kuδ'mɜ** "Asche" (UEW 194)
A: Tk ***kül** "Asche" (VEWT 307) | MKor ***kɨ́rí-m** "soot" (AED).

30. ***kuma**"black / coal"
AA: Sem: Aramaic **ukkāmā** id. (Calice) | Eg **km** "black"
U: FW ***kūma** "heiss, glühend; Fieber" (UEW 675-76)
D ***kam-** "be singed, burnt" (DEDR 1230)
A ***k(i̯)ume**: Tk ***kömür** "coal" (VEWT 289) | OKor (Koguryo) **kəmər** "black" (Ki-Moon Lee 1977, 38), MKor ***kə̄m-** "black" (AED).

31. ***k‘awk/ga** "high / long"
IE ***kouko-**: Gmc ***háuha-/*haugá-** "high" (Kluge 1975, 312), maybe Toch A **koc**, B **kauc** id. (Adams 1999, 209)
U: FU ***kawka** "long" (UEW 132).

32. ***k’ibʌ** "cold"
AA: Eg **qb(b)** "cold" | Cush: (C) ***kaNb-** "cold" | (E) ***qab-** "cold" (cf. Dolgopolsky 1983, 135) | Om: (N) Kafa **kabo** "fever", Shinasha **kuppo** id. | ?Ch: (E) Kwang **kabár** "wind" (Lukas)
D ***kīv-** "cold, cool" (DEDR 1618).

33. ***k‘oni(dʌ)**? "belly"
AA: Eg **qnj** "bosom, embrace" (Wb. V, 50), Demotic **qn** "lap", Copt **kun** "bosom" (Černý 1976, 59) | Om: Dache **ḳance**, Sezo **kaani** "belly" (Bender) | Ch: (W) Hausa **ḳundu** "bird's stomach", Ngizim **kwùnú**, Bade **kúnà-an** id.
U: FU ***konз** "belly" (UEW 208)
?D ***kuṇṭ-** "heart, kidney, chest" (DEDR 1693).

34. ***k‘on(y)ʌ** "back"
AA ***ḳayn- (*ḳawin-?)**: Sem: Akk **qinnatu** "Hinterbacke", Arab **qaynat** "Hinterteil", Tigre **qen** "unterer oder hinterer Teil" (Christian, *WZKM* 1959, 168-70) | Cush: (E) Oromo **qinṭii** "spine" (Gragg) | Ch: (W) Hausa **ḳaniya** "anus"; (C) Gava **kwinda**, Bokwa **kunja**, Gboko **kunj** "Nacken" (Büchner)
D: Tamil **kuṇṭi** "buttocks", Kuwi **kūna** "buttock" (DEDR 1693A).

35. ***k‘uma ~ *k‘amu** "night / cloud"
AA: Cush: (C) Awngi **kemana** "evening" (Conti Rossini) | Om: (N) Wolayta **ḳama**, Kullo **ḳamma**, Koyra **ḳaama**, Gidicho **ḳaammani** "night" (Bender) | Ch: (C) Wandala, Hurzo **kumba** "cloud" (Mouchet)
IE: Hit **kammara-** "Rauch, Dunst, Qualm, Wolke", Gr κέμμερος · ?χλύς (B. Čop, *Indouralica I*, 1974, 99)
U: FU ***kumз** "Wolke" (UEW 204-05),
A: MKor **kɨmɨ-** "become dim, hide (of moon)" | OJp **kumor-** id., cf. pJp ***kùmua-N** "cloud" (AED).

36. ***liš(H?)a** "meat / flesh"
Kartv: Georgian **leši** "flesh, meat"
U: BF ***liša**?: Fi **liha** "meat, flesh", Lappish T **liške** "body, corpse" (SKES 292).

37. ***l’owk’ʌ** "lake, (reservoir of) water"
AA: Om: (S) Ubamer etc. **luḳa / loγa**, Galila **lu·ga / lo·ga** "water" (Fleming), Banna **nɔ:qɔ**, Dime **nə:γo** id. (Bender) | Ch: (C) Kotoko ***lVγVm-** "river" (Porxomovskij)
IE ***laku- ~ *loku-** "lake" (P 653)
U ***l’owkkз** "Loch, Grube" (UEW 252; Čop, *ALH* 24, 1974, 98-99: IE + U)

38. ***muk**$^{(‘)}$**a ~ *mak**$^{(‘)}$**u** "neck"
AA ***m[u]k-**: Eg **mkḥ3** "Hinterkopf" (Wb. II, 163), cf. **ḥ3** "nape" | Cush: (N) Beja **mook** "neck" (Reinisch) | (C) Bilin **məkk**^{w}**a** "buttocks" | (E) Afar **makuḥ**, **mukoḥ** "spine, spinal cord" (Parker-Hayward), Boni **múkkə** "buttocks" (Heine), Yaaku **muk** "lower side of body" (Ehret)
?IE: Iranian of Pamir: Shugni, Wakhi **māk**, Sarikoli **mok**, Ishkashim **mak** "back of the neck, nape" (Morgenstierne) - a Dravidian-like substratal influence?
U ***muka** "Rücken" (Janhunen)
D ***mak(k)-** "neck" (DEDR 4622)
A ***mi̯ák‘ù** "neck": MTk **bakan** "Halsband" (VEWT 58) | Tg: Evenki **muka** "skin of the deer’s neck" (TMS I, 551) | MKor ***mòk** "neck, throat" | pJ ***múk-** (AED).

39. ***naKuŕi** "river / lake"
D ***nākuṳ̄ri**
A· Mong **naγur**, Kalm **nūr** "See, Teich" (KW 282) | OKor (Silla) ***narih**, MKor **nayh** "river" (Ki-Moon, Lee 1977, 80).

40. ***ńaǯi(r)** "sun / midday / summer"
D ***ñāciṟ-** (DEDR) = ***ńēsiṟ** (G. Starostin) "sun" (DEDR 2910)
A ***nāǯi(-r?)** "midday / summer": Tk ***jāj** "spring, summer" (VEWT 179) | Mong ***naǯir** "summer" | MKor ***náč** (AED).

41. ***p/biša** "green / bile"
AA: Cush: (E) ***bis-** "color; flower" (Sasse 1982, 37)
IE ***bis-(t)li-** "bile" (P 102)
U ***piša** "Galle; grün, gelb" (UEW 384-85)
D ***pac-** "green", cf. Kolami **payt** "bile" (DEDR 3821)
Note: Dolgopolskij, *Etimologija* 1964[65], 266: IE + U; Tyler, *Language* 1968, 807: U + Dr.

42. ***p**$^{(ʿ)}$**olʌ** "flow"
IE ***pleu-** "to flow" (P 835-36)
U: Ugr ***pɤlз** "quellend fliessen" (UEW 881)
D ***pol-** "to was" (DEDR 4549) or ***puḷ-** "(to) float" (DEDR 4321)
ChuK: Chu **pylyl** "rasche Strömung; fliessen, strömen" (Bouda, *Lingua* 4, 1954-55, 301: Chu+Hu)

43. ***p**$^{(ʿ)}$**utʌ** "bird / feather"
IE: Balto-Slavic ***put-**: Lithuanian **putýtis** "young animal/bird", Latvian **putns** "bird"; Church Slavonic **pъta** id. (VT III, 398)
D ***put-** "feather" (DEDR 4278).

44. ***pʌnʌ(-Cʌ)** "louse; mosquito"
AA: Ch: (C) Chibak **pinzù**, Bura, Ngwaxi **pinjù**, WMargi **pınju** "mosquito" (Kraft)
U: Mordvinian **panǯam** "ant"; Nenets **pansie**, **panźe** "louse" (Schrader, *Zeitschrift für Indo-Iranistik* 3, 1925, 93: U + D)
D ***pēn** "louse" (DEDR 4449).

45. ***p'elʌ** "ashes"
IE ***pel-(en-)** "ashes" (P 802)
U: FP ***pelme/*pel'me** "Schmutz, Staub, Asche" (UEW 728)

46. ***p'idʌ** "long"
AA: Om: (N) Bench **ṗaad**, She **ṗad** "long" (Bender)
U ***piδe** & ***piδke** "hoch; lang" (UEW 377)
D ***poṭ-** "long, tall, high" (DEDR 4484).

47. ***p'ojamu** "snake"
?AA: Ch: (W) Sura **púpwáp**; Fyer **púpwáp** "fish" (JgIb 140)
D ***pām-** "snake" (SD ***pāmb-**) (G. Starostin)
A ***p'[ò]jamV** "snake, worm": Tk ***uman**: Chuvash **ъ**^{w}**man** "worm" | *Mong **jamu** "worm" (KW 214) | Tg ***püjmur** "dragon, monster" (TMS I, 466) | MKor ***pă´jàm** "snake" | pJp ***pàimpV** "snake" (AED).

48. ***p‘öli** "louse / flea / fly"
AA: Sem: Arab **falā** "to remove lice", **fāliyat** "sorte d`insecte semblable au scarabée", Harsusi **felō** "to delouse"; cf. also Akk **uplu** "Kopflause" | Eg **pj** "flea" | (C) NAgaw ***filut-/fälät-/fələt-** "flea"; (E) Dullay ***fillay-** "flea" (AMS) | Om: (N) Kafa **p/ṗillo** "flea, bedbug" (Reinisch), Wombera **pela** "locust" (Fleming) | Ch (W) Bokkos **ḅwele** "lice"; (C) Gulfei **fe̥lei̯**, Musgu **áfili**, Mbara **fùlay** "mosquito" (EDE II, 411-12)
U **pälä** "horsefly" (UEW 416)
A ***p‘i̯oli** "fly, midge": Mong ***hilaγan** (KW 206) | Tg ***pulmi-kte** (TMS II, 348) | MKor ***pă´rh** (AED).

49. ***p’ulʌ** "grass"
D ***pul** "grass" (DEDR 4300)
A ***p‘ōle**: Tk ***ȫl** "wet" (VEWT 371) | Tg ***pul-**: Even **hul** (bot.) "horse-tail" (TMS II, 342) | MKor **pï´rh** "grass", OKor (Koguryo) ***pă(l)lŏl** "green" (AED)

50. ***p’ülʌ** "ashes / dust"
AA: Ch: (W) NBauchi ***ṗuli** "ashes" > Miya **pə̀lí**, Warja **pwə̀lèèná**, etc. (cf. Stolbova 1987, 248; JgIb 4)
IE ***pelw-** "dust" (P 802)
U: BF: Fi **pöly** "dust", Vepsian **pölü** id. (SKES 696)
D: Tamil **pūḻi** "dust, holy ashes", Tulu **poyye** "sand" (Mayrhofer II, 512)
A ***p/p‘ülńe** "ashes": Mong ***hüne-sü** id. (KW 458) | Tg ***pulńe-** id. (TMS II, 347; reconstructions after AED)

51 ***sawina** "know"
AA: Sem: Hbr **šənīnah** "something which is knowing or told everybody", Arab **sunna** "tradition" | Berb **√w-s-n**: Tuareg of Ahaggar **essen**, caus. **s-ussen** "to know" | Eg **swn** "to recognize", later "to know", Copt **sōwn** "to know" (Černý 1976, 168) | Ch: (W) ***sayn-** "to know" (Stolbova 1987, 178); (C) Margi **səni**; Mafa **sun, sin**, Gidar **sən**; Logone **sən**, Ngala **seŋ**, Yedina **hin** id. (Greenberg 1963, 58)
IE ***swen-** "tönen" (P 1046), cf. Persian **xonjā** "Musik, Melodie, Gesang", **xvāndan** "rufen, lesen", Old Irish **son** "Laut", Welsh **hanes** "history", Old English **swinn** "Musik, Gesang", OIc **svinna** "wisdom" (Mann 1984-87, 1346)
A. Mong ***sonos-** "to hear" (KW 331).

52. ***Sir(j)u** ~ ***Sürjä** "(top of) hill, mountain"
?AA: Cush: (E) Saho **saray,** Som **saray** "Oberseite, Höhe" (Reinisch)
Kartv: Georgian **seri** "hill", Laz **serṭi** "top of the hill" (cf. Tailleur, *Lingua* 12, 1963, 401)
IE: Hittite **sār** "above", Cuneiform Luwian **sarri**, Lycian **hri** id.; ?Greek ‘ρίον "Berghöhe, Vorgebirge", Mycenaean **ri-jo** indicating ***sriyo-** and not ***wriyo-** (Heubeck, *Orbis* 13, 1964, 266)
U: FV ***śürjä** "Seite, Kante, Rand" (UEW 779)
D ***cōr-** "hill, mountain" (DEDR 2887)
A ***sira/u**: Tk ***sɨrt** "mountain ridge" (VEWT 419) | Mong ***siru/o** | Tg ***sirk-**: Evenki **sirkun** "cape" (TMS II, 95, AED).

53 ***tiŋka(lʌ)** "moon"
IE ***dn̥g**^{h}**-(lo-)**: Gmc ***tungla-** n.: OIc **tungl** "moon", **himin-tungl** "constellation", OSax **tungal**, OHG **himil-zungal** "constellation", Goth **tuggl** id., maybe comparable with Lith **diñga man** "es dünkt mich" (de Vries 1977, 601)
D ***tiṅkaḷ** "moon" (DEDR 3213).

54. ***t‘arwi** ~ ***t‘awri** "leg"
AA: Berb: Siwa **ṭår**, Ghadames **aḍar**, Iznacen **ḍār**, Zenaga **əḏa’r**^{h} , Tuareg of Ahaggar **aḍər** "leg, foot" (A. Militarev 1991, 260-61) | ?Om: (S) Karo, Banna **rro**, Dime **dɔɔto** "foot" (Fleming & Bender)

Kartv: Georgian **ṭor-** "paw"
D ***tār̤** ~ ***tāḷ** "leg, stem" (DEDR)
A ***t'i̯ūŕe**: Tk ***diŕ** "knee" (VEWT 482) | Mong ***türeji** "Stiefelschaft" (KW 415) | Tg ***türē-kse** id. (TMS II, 238-39) | MKor ***tàrí** "leg" (AED).

55. ***t'igu(nʌ)** "hear"
AA ***ṭig^{w}-(n-)**: Cush (E) ***ḍeg-/ *ḍog-** "to hear; ear" (Sasse 1982, 61) | Ch: (W) NBauchi ***n-dukw/y-**; Bade **dúgwâ** "to hear" (Stolbova 1987, 252); (C) Tera **təkì** id.; Masa **togor** "ear"; (E) Kwang **tugum** "ear", Dangla **ḍéngei**, Jegu **'údúŋê** id., **ḍoj-**, Sumray **doy**, Mubi **jige**, Lele **ḍeeŋli** "to hear" (Dolgopolskij 1973, 55, 257; JgIb 184-85).
A ***t'iŋ-**: Tk ***tïŋ-** "to hear, listen" | Mong ***čiŋna-** "to listen" (VEWT 478; Sevortjan III, 236-37).

56. ***t'ut'i** "ashes / smoke"
AA: Ch: (W) Dera **duḍa** "ashes"; Ron: Kulere **'a-totə** id. (Jungraithmayr)
Kartv ***ṭuṭa-** "ashes" (Klimov 1964, 183)
A ***t'ut'i** "smoke": Mong ***tüt-ün, -süg** | MK **tɨtkɨr/ tɨthɨr** (AED).

57. ***wejCʌ** "all"
AA: Om: Dokko (Fleming) **wayci**, Basketo (Bender) **wo̲ica** "all"
IE: Balto-Slavic ***wiso-**: Lith **vìsas** "ganz", Latv **viss** "ganz, all(es)", Pruss **wissa** "all(erlei)"; Sl ***vьsъ** "all" (VT I, 304-05)
U ***we(ń)ćɛ** "ganz, all" (UEW 568)
D ***vic-** "all" (DEDR 5387).

Abbreviations

A Altaic; AA Afro-Asiatic; Akk Akkadian; Arab Arabic; Aram Aramaic, Arm Armenian; Berb Berber; BF Balto-Fennic; Ch Chadic; Chu Chukcki; ChuK Chukchi-Kamchatkan; Copt Coptic; Cush; D Dravidian; E East; Eg Egyptian; ESA Epigraphic South Arabian; Esk Eskimo; Fi Finnish; FP Fenno-Permian; FU Fenno-Ugric; Gmc Germanic; Gr Greek; Hbr Hebrew; Hit Hittite; Hu Hungarian; Ic Icelandic; IE Indo-European; Kalm Kalmyk; Kartv Kartvelian; Kor Korean; Latv Latvian; Lith Lithuanian; M Middle; Mong Mongolian; N North; Om Omotic; Pruss Prussian; S South; Sem Semitic; Sl Slavic; Som Somali; Syr Syriac; Tg Tung; Tk Turkic; U Uralic; Ug Ugaritic; W West; Wr Written.

References

Adams, D.Q. 1999. *A Dictionary of Tocharian B*. Amsterdam - Atlanta: Rodopi.
AED = *Altaic Etymological Dictionary*, by Starostin, S., Dybo, A. & Mudrak, O. Moscow 1995: Ms.
AHw von Soden, W. 1965-81: *Akkadisches Handwörterbuch*. Wiesbaden: Harrassowitz.
AMS = Amborn, H., Minker, G. & Sasse, H.-J. 1981. *Das Dullay*. Berlin: Reimer.
Appleyard, D.L. 1984: The internal classification of the Agaw languages: A comparative and historical phonology. In: J. Bynon (ed.), *Current Progress in Afro-Asiatic Linguistics. Papers of the Third International Hamito-Semitic Congress*. Amsterdam - Philadelphia: Benjamins, pp. 33-67.
Bender, M.L 1971: The languages of Ethiopia. A new lexicostatistic classification and some problems of diffusion. *Anthropological Linguistics* 13/5, pp. 165-288.
Bender, M.L. 1975: *Omotic: A New Afroasiatic Language Family*. Carbondale: University Museum Studies.
Bender, M.L. 1994: Aroid (South Omotic) Lexicon. *Afrikanistische Arbeitspapiere* 38, pp. 133-162.
Biella, J.C. 1982: *Dictionary of Old South Arabic*. Chico: Scholar Press.
Black, P. 1974: *Lowland East Cushitic: Subgrouping and Reconstruction*. Yale University Ph.D. Dissertation.
Blažek, V. 1989-90: Lexica Nostratica: Addenda et Corrigenda, I-II. *Archiv orientální* 57, pp. 201-10 & 58, pp. 205-218.

Blažek, V. 1992: The new Dravidian-Afroasiatic parallels. Preliminary report. In: *Nostratic, Dene-Caucasian, Austric and Amerind*, ed. V. Shevoroshkin. Bochum: Brockmeyer, pp. 150-65.

Blažek, V. 2002: Some New Dravidian-Afroasiatic Parallels. *Mother Tongue* 7, 171-198.

Boisson, C. 1989[ms.]: *Sumerian / Nostratic / Sino-Caucasian isoglosses*. Lyon: Ms.

Bomhard, A.R. 1984: *Toward Proto-Nostratic*. Amsterdam - Philadelphia: Benjamins.

Bomhard, A.R. & Kerns, J.C. 1995: *The Nostratic Macrofamily: A Study in Distant Linguistic Relationship*. Berlin - New York - Amsterdam: Mouton de Gruyter.

BSOAS = *Bulletin of the School of Oriental and African Studies*.

Calice, F. 1936: *Grundlagen der ägyptisch-semitischen Worvergleichung*. Wien Oriental. Institut des Uni

Cerulli, E. 1936-51: *Studi etiopici*, I-IV. Roma: Istituto per l'Oriente.

Cohen, D. et al. 1970f: *Dictionnaire des racines sémitiques*. Paris-La Haye: Mouton.

Cohen, M. 1947: *Essai comparatif sur le vocabulaire et la phonétique du chamito-sémitique*. Paris: Champignon.

Černý, Jaroslav. 1976: *Etymological Dictionary of Coptic*. Cambridge: University Press.

DEDR = Burrow, T. & Emeneau, M. 1984: *A Dravidian Etymological Dictionary*, 2nd ed. Oxford Clarendon Press.

Dolgopolsky [Dolgopol'skij], A. 1966: Materialy po sravnitel'no-istoričeskoj fonetike kušitskix jazykov: gubnye i dental'nye v načal'nom položenii. In: *Jazyki Afriki*, ed. B.A. Uspenskij. Moskva: Nauka, pp. 35-88.

Dolgopolsky [Dolgopol'skij], A. 1973: *Sravnitel'no-istoričeskaja fonetika kušitskix jazykov*. Moskva: Nauka.

Dolgopolsky, A. 1983: Semitic and East Cushitic: Sound correspondences and cognate sets. In: *Ethiopian Studies (Fs. W. Leslau)*, eds. S. Segert & A J.E. Bodrogligeti. Wiesbaden: Harrassowitz, pp. 123-142.

Dybo, A. 1988: Étimologičeskij material k rekonstrukcii pratungusomaňčžurskix nazvanij častij tela. In: *Sinxronija i diaxronija v lingvističeskix issledovanij*, I. Moskva: Institut vostokovedenija, pp. 108-27.

Dybo, A. 1989: Methods in systemic reconstruction of Altaic and Nostratic lexics. In: *Lingvističeskaja rekonstrukcija i drevnejšaja istorija Vostoka*, I. Moskva: Nauka, pp. 196-207.

Dybo, A. 1996: *Semantičeskaja rekonstrukcija v altajskoj ėtimologii: somatičeskie terminy (plečevoj pojas)*. Moskva: Studia Philologica

EDE = Takács, Gábor. 1999-2001: *Etymological Dictionary of Egyptian*, Vols I-II. Leiden - Boston - Köln. Brill.

Ehret, C. 1980: *The Historical Reconstruction of Southern Cushitic Phonology and Vocabulary*. Berlin: Reimer.

ESSJ = Ėtimologičeskij slovaŕ slavjanskix jazykov, ed. by O.N. Trubačev. Moskva: Nauka 1974f.

Fähnrich, H. & Sardshweladse, S. 1995: *Etymologisches Wörterbuch der Kartwel-Sprachen*. Leiden-New York-Köln: Brill.

Gragg, G. 1982: *Oromo Dictionary*. East Lancing African Studies Center.

Greenberg, J.H. 1963. *Languages in Africa*. Bloomington: Indiana University.

Haberland, E. & Lamberti, M. 1988: *Ibaaddo ka-Ba'iso: culture and language of the Ba'iso*. Heidelberg: Winter.

Hayward, D. 1979: Bayso revisited. *BSOAS* 44, pp. 126-144.

Hayward, D. 1984: *The Arbore Language*. Hamburg: Buske.

Heine, B. 1978: The Sam languages. A history of Rendille, Boni and Somali. *Afroasiatic Linguistics* 6/2, pp. 23-115.

Heine, B. 1981: Some cultural evidence on the early Sam-speaking people of Eastern Africa. *Sprache und Geschichte in Afrika* 3, pp. 169-200.

Heine, B. 1982: *Language and Dialect Atlas of Kenya: Boni Dialects*. Berlin. Reimer.

Hetzron, R. 1978: The nominal system of Awngi (Southern Agaw). *BSOAS* 41, pp 121-141.

Hudson, G. 1989: *Highland East Cushitic Dictionary*. Hamburg Buske.

Illič-Svityč, V.M. 1971-76-84: *Opyt sravnenija nostratičeskix jazykov*, I-III Moskva: Nauka.

JgIb = Jungraithmayr, H. & Ibriszimow, D 1994: *Chadic lexical roots*, Vol II: *Tentative reconstruction, grading and distribution*. Berlin. Reimer.

KESK = Lytkin, V.I. & Guljaev, E.I. 1970: *Kratkij ėtimologičeskij slovaŕ komi jazyka*. Moskva: Nauka.

Ki-Moon, Lee. 1977: *Geschichte der Koreanische Sprache*. Wiesbaden: Reichert.

Klein, E. 1987: *A Comprehensive Etymological Dictionary of the Hebrew Language*. New York - London: Macmillan.

Klimov, G.A. 1964: *Étimologičeskij slovaŕ kartvel'skix jazykov*. Moskva: Nauka.

Kluge, F. 1975: *Etymologisches Wörterbuch der deutschen Sprache*$_{21}$. Berlin-New York: Walter de Gruyter.

Koskinen, K.E. 1980: *Nilal. Über die Urverwandtschaft des Hamito-Semitischen, Indogermanischen, Uralischen und Altaischen*. tampere: Author.

Kraft, C. 1981: *Chadic Wordlists*, I-III. Berlin: Reimer.

KW = Ramstedt, G.J. 1935: *Kalmückisches Wörterbuch*. Helsinki: Lexica Societatis Fenno-Ugricae III.

Lamberti, M. 1986: *Die Somali-Dialekte*. Hamburg: Buske.

Lamberti, M. 1987: Some Konsoid Etymologies. *Anthropos* 82, pp. 529-541.

Lamberti, M. 1993: *Die Shinasha-Sprache*. Heidelberg: Winter

Lamberti, M. & Sottile, R. 1997: *The Wolayta Language*. Köln: Köppe.

Leslau, W. 1987: *Comparative Dictionary of Ge^{c}ez (Classical Ethiopic)*. Wiesbaden: Harrassowitz.

Leus, T., Van de Loo, J. & Cotter, G. 1992: *A Vocabulary of Oromo- English*. Debre Zeït (Ethiopia).

Lukas, J. 1937· *Zentralsudanische Studien*. Hamburg: Friedrichsen, de Gruyter & Co.

M = Mann, S.E. 1984-87: *An Indo-European Comparative Dictionary*. Hamburg: Buske.

Mayrhofer, M 1986-92: *Etymologisches Wörterbuch des Altindoarischen*, I-II. Heidelberg: Winter.

Militarev, A. 1991: Istoričeskaja fonetika i leksika livijsko-guančskix jazykov In: *Jazyki Azii i Afriki*, IV (Afrazijskie jazyki) 2. Moskva: Nauka, pp. 238-267.

Mouchet, J. 1950: Vocabulaires comparatifs de quinze parlers du Nord-Cameroun. *Études camerounaises* III/29-30, pp. 5-74.

Mouchet, J. 1953: Vocabulaires comparatifs de sept parlers du Nord-Cameroun. *Études camerounaises* VI/41-42, pp. 137-199.

Mukarovsky, H.G. 1987: *Mande-Chadic. Common Stock. A study of Phonological and Lexical Evidence*. Wien: Afro-Pub (Beiträge zur Afrikanistik, Bd. 32).

Parker, E.M. & Hayward, R.J. 1985. *An Afar-English-French Dictionary*. London: SOAS.

P = Pokorny, J. 1959*: Indogermanisches etymologisches Wörterbuch*. Bern-München: Francke.

Prasse, K.-G. 1972-74: *Manuel de grammaire touarege* (tăhăggart), I-III, IV-V, VI-VII. Copenhague Akademisk Forlag.

Rabin, Ch. 1975: Lexicostatistics and the internal divisions of Semitic. In· *Hamito-Semitica (Proceedings of a Colloquium held in London* 1970), eds J. & T Bynons . The Hague-Paris Mouton, pp. 85-99.

Reinisch, L. 1887: *Die Bilin-Sprache in Nordost-Afrka*. Wien Hölder.

Reinisch, L. 1895: *Wörterbuch der Beḍauye-Sprache*. Wien: Hölder.

Reinisch, L. 1902: *Die Somali-Sprache*, II: *Wörterbuch*. Wien· Hölder.

Roper, E.-M. 1928: *TūBeḍawiạ*. Hertford: Herts.

Sasse, H.-J. 1979: The consonant phonemes of Proto-East Cushitic (PEC) *Afroasiatic Linguistics* 7/1, pp. 1-67.

Sasse, H.-J. 1982: *An Etymological Dictionary of Burji*. Hamburg: Buske.

SED = Militarev, A. & Kogan, L. 2000: *Semitic Etymological Dictionary*. Münster: Ugarit-Verlag.

Sevortjan, G. 1974f: *Étimologičeskij slovaŕ tjurkskix jazykov*. Moskva: Nauka

SISAJ = *Sravnitel'no-istoričeskij slovaŕ afrazijskix jazykov*, I-III, eds. A.G. Belova, I.M. D'jakonov, A.Ju Militarev, V.Ja. Porxomovskij, O.V. Stolbova, A.C. Četveruxin. Moskva: Nauka 1981, 1982, 1986 (*Pis'mennye pamjatniki i problemy istorii i kul'tury narodov Vostoka*, XV/IV, XVI/III & XIX/III *godičnaja sessija LO IV AN SSSR*).

SKES = Toivonen, Y.H. et al. 1955f: *Suomen kielen etymologinen sanakirja*. Helsinki: Lexica Societatis Fenno-Ugricae XII, 1f.

Skinner, N. 1977: The Northern Bauchi Chadic Languages. *Afroasiatic Linguistics* 4/1, pp. 1-49.

Skinner, N. 1996: *Hausa Comparative Lexicon*. Köln: Köppe.

Sölken, H. 1967: *Seetzens Åffadéh. Ein Beitrag zur Kọtoko-Sprachdokumentation*. Berlin: Akademie-Verlag

Starostin, S.A. 1991: *Altajskaja problema i proisxoždenie japonskogo jazyka* Moskva: Nauka.

Steingass, F. 1988: *Arabic-English Dictionary*. Jalandhar City. Gaurav

Stolbova, O.V. 1987: Sravnitel'no-istoričeskaja fonetika i slovaŕ zapadnočadskix jazykov. In. *Afrikanskoe istoričeskoe jazykoznanie*, ed. V. Porxomovskij. Moskva Nauka, pp 30-268.

Stolbova, O.V. 1996: *Studies in Chadic Comparative Phonology* Moscow. Diaphragma.

TMS = *Sravnitel'nyj slovaŕ tunguso-mańčžurskix jazykov* I-II, ed. V.I. Cincius et al. Leningrad: Nauka 1975-77.

Tosco, M. 1991: *A grammatical sketch of Dahalo*. Hamburg: Buske.

Tourneux, H., Seignobos, Ch. & Lafarge, F. 1986: *Les Mbara et leur lange (Tchad)*. Paris: SELAF.

UEW Rédei, K. (ed.) 1986-88: *Uralisches etymologisches Wörterbuch*. Budapest: Akadémiai Kiadó.

Van Windekens, A.J. 1976: *Le tokharien confronté avec les autres langues indo-européennes, I: La phonétique et la vocabulaire*. Louvain: Centre International de Dialectologie Générale.

VEWT = Räsänen, M. 1969: *Versuch eines etymologischen Wörterbuchs der Türksprachen*. Helsinki: Lexica Societatis Fenno-Ugricae XVII, 1.

VT = Vasmer, Max. 1986-88. *Étimologičeskij slovaŕ russkogo jazyka*, I-IV. Moskva: Progress (from the German original *Etymologisches Wörterbuch der russischen Sprache*, Heidelberg: Winter 1950-58, translated by O.N. Trubačev)

Vycichl, W. 1983: *Dictionnaire étymologique de la langue copte*. Leuven-Paris. Peeters.

Wb. = Erman, A. & Grapow, W 1928-31[1971]: *Wörterbuch der ägyptischen Sprache*, I-VI. Berlin. Akademie-Verlag.

Václav Blažek
blazek@phil.muni.cz

Notes on Basque Comparative Phonology[1]

John D. Bengtson
Association for the Study of Language in Prehistory,
Santa Fe Institute

The genetic classification of the mysterious Basque language has been a topic of vigorous discussion throughout the Twentieth Century, and this discussion continues to the present day. Some of the latest exchanges in this discussion are found in the journals *Mother Tongue* (especially issues I and V) and *Dhumbadji!* (see References). Impelled both by justified criticism and encouragement (see, *e.g.*, Blažek 1995, Starostin 1996), I have worked intermittently for several years at establishing regular phonological correspondences between Basque and the languages that are most closely related to it,[2] namely the (North) Caucasian languages, and Burushaski. In several articles published previously I have used the terms "Macro-Caucasian" or "Vasco-Caucasian" for this hypothetical language family.[3] In the present paper comparisons will primarily be made within the Macro-Caucasian family (Basque + Caucasian + Burushaski). Lexical comparisons involving only Caucasian and Burushaski (and not Basque) will be set aside for the time being.

The Sound System of Basque

A common Basque consonant inventory is as follows (modified from Hualde, 1991):

labial	**dental/alveolar**[4]		**palatal**	**velar**
p	**t**		**c**	**k**
b	**d**		**ɉ**	**g**
f	**s̄**	**ś**	**š**	**x**
	c̄	**ć**	**č**	
m	**n**		**ñ**	
	l		**ʎ**	
	r			
	r̄			

In standard Basque orthography the phonemes **/p, t, k, b, d, g, f, m, n, ñ, l, r/** are written with the corresponding Latin or Spanish letters: *p, t, k, b, d, g, f, m, n, ñ, l, r*. The remaining phonemes are represented as follows:

---

[1] This paper consists of excerpts from a much longer paper written for the Workshop: "Basque Phonology in a Dene-Caucasian Context" (75 pp).

[2] Based on lexical and morphological evidence, only a small part of which is presented here.

[3] I consider Macro-Caucasian to be a sub-group of the larger family (macro-family) "Dene-Caucasian" (Starostin's "Sino-Caucasian") that also includes the Sino-Tibetan, Yeniseian, and Na-Dene families. There is substantial evidence that the Salish and Wakashan language families may also be part of Dene-Caucasian. (See Vitaly Shevoroshkin's article, "Salishan and North Caucasian," in this issue.)

[4] **/t, d/** are dental; **/l, n, r, r̄/** are alveolar; **/s̄, c̄/** are dorso-alveolar (= lamino-alveolar); **/ś, ć/** are apico-alveolar.

c = *tt* as in *ttipi* 'very small, tiny' (dim. of *tipi*)
ɉ̇ = *dd* as in *onddo* 'toadstool, fungus'
s̄ = *z* as in *zazpi* 'seven'
c̄ = *tz* as in *hotz* 'cold'
ś = *s* as in *sei* 'six'
ć = *ts* as in *huts* 'fault, empty'
š = *x* as in *xexen* 'small bull' (dim. of *zezen*)
č = *tx* as in *txitxar* 'grasshopper, cicada'
ʎ = *ll* as in *bello* 'hot' (dim. of *bero*)
r̄ = *rr* as in *harri* 'stone'
x = *j* as in *jan* 'to eat' (but only in western Basque, mainly Gipuzkoa)

Elsewhere the letter *j* is pronounced **[dž]** (Bizkaia), **[j]** (Alto Navarro), **[ɉ]** (Lapurdi, Basse Navarre), **[ž]** (Zuberoa), or even other variants.[5]

Northern ("French" Basque) dialects (Lapurdi, Basse Navarre, Zuberoa) also have a phoneme **/h/**, generally corresponding to orthographic *h*. These same dialects lack the phoneme **/x/**, and also tend to have aspirated consonants, such as Zuberoan *phiper* 'pepper' (but BN, L *bipher*), *thu* 'spit', *khedarre* 'soot', *anhua* 'provisions', *iñhar* 'a little', *alhaba* 'daughter', *ürhe* 'gold', etc.

The voiced obstruents written *b, d, g* have stop **[b, d, g]** or fricative **[β, ð, γ]** allophones, depending on their position in the word or phrase.[6] Trask (1997) prefers to call the latter "continuants" (approximants with no audible friction).

The Basque vowel system is a simple 5-vowel system: **/a, e, i, o, u/**. Only the Zuberoan (Souletin) dialect differs in also having the front-round vowel **/ü/**. Zuberoan and Roncalese also have a contrastive set of nasal vowels **/ã, ẽ, ĩ, õ, ũ, (ü̃)/**.

The Sound System of Proto-Caucasian

In this paper the Proto-Caucasian (PNC) reconstruction by Nikolayev and Starostin is accepted as a baseline, while recognizing that some details are open to correction or modification. Nikolayev and Starostin (1994: 40) postulate the following consonant phonemes:[7]

	Occlusives			**Fricatives**		**Resonants**		**Glides**
	voiceless	**voiced**	**glottalized**	**voiceless**	**voiced**			
labial	p	b	ṗ	f		w	m	u̯
dental	t	d	ṭ			r	n	j
hissing	c	ʒ	c̣	s	z			
palatal (hissing-hushing)	ć	ʒ́	ć̣	ś	ź			

---

[5] See the map presented by Trask (1997, p. 86).

[6] This pattern is similar to, but not identical with, the pattern in Castilian Spanish. (See Hualde 1991.)

[7] "Two more very rare voiced fricatives are reconstructed for PEC (lateral L and velar ɣ), as well as the supposedly interdental fricatives ϑ and ϑ:. These phonemes have no correspondences in PWC, and their existence in PNC is dubious." (NCED, p. 41)

hushing	č	ǯ	č̣	š	ž			
lateral	ƛ	Ł	ƛ̣	λ		l		ɫ
velar	k	g	ḳ	x				
uvular	q	G	q̇	χ	ʁ			
laryngeal			ʔ	h	ɦ			
emphatic laryngeal			ʡ	ħ	ʕ			

(For explanations of the phonetic symbols see Appendix A.)

Nikolayev and Starostin (NCED, p. 72) reconstruct 9 vowels for Proto-Caucasian, each of which may be short or long:

i	ü	ɨ	u
e		ə	o
ä		a	

1. Vowel Correspondences between Basque and Proto-Caucasian

Although some details remain to be resolved, numerous examples verify the usual correspondences.

Basque /**a**/ regularly corresponds to Caucasian _*a (*ă, *ā):_

1.1. Basque _(h)a-, -a_ (article, demonstrative) ~ PEC _*hă_ demonstrative stem (NC 436)
1.2. Basque _sare, sale_ 'net, grate; stockade' ~ PEC _*čɦałē_ 'enclosure, fence' (NC 343)
1.3. Basque _hatz_ 'finger, paw', _be-hatz_ 'toe, thumb' ~ Avar _k_^{W}_ač̣_ 'paw', etc. < PNC _*kwăn Vč̣ĕ_ (NC 704)
1.4. Basque (B) _apo_ 'hoof' ~ Bezhta, Hunzib _ap'a_ 'paw', etc. < PNC _*HaṗV̆_ (NC 545)
1.5. Basque (BN, Sal) _udagara_ 'otter' ~ PEC _*darq̇wV_ > Andi _dar_ɢ^{w}_a_ 'weasel, marten', etc. (NC 399)
1.6. Basque _izar_ 'star', etc. ~ PNC _*ʒwăhrī_ 'star' > Tindi _c:aru,_ etc. (NC 1098)
1.7. Basque _sabel_ 'belly, stomach' ~ PNC _*ǯăbV_ 'kidney, liver' (NC 1106)
1.8. Basque _lasto_ 'straw' ~ PEC _*ƛ̣ačă_ 'leaf; a kind of plant' (NC 773)
1.9. Basque _lamika_-tu, _lamizka_-tu 'to lick' ~ Andi _lam-_ 'to lick', etc. < PEC _*łam V_ (NC 754)
1.10. Basque _sagu_ 'mouse' ~ PNC _*cārgwɨ̄_ 'weasel, marten, mouse' (NC 322)
1.11. Basque (B, G) _apal_ 'shelf' ~ PEC _*ʔap VłV_ 'pole; board, cover' (NC 202)
1.12. Basque _hari, hal-_ 'thread' ~ PEC _*χ̱āłV_ 'sinew, thread' > Chechen _χ̱al_, etc. (NC 1067)
1.13. Basque _handi_ 'big, great' ~ Proto-Circassian _*k_^{hw}_a(n)də_ 'much, many' (Kuipers 1975)
1.14. Basque _tak-oi(n)_ 'heel (of a shoe)' ~ PNC _*dHāq̱wĀ_ 'back of head', etc. (NC 399)
1.15. Basque (R) _atze_ 'tree'~ PNC _*Hă(r)ǯwī_ 'tree' > Khwarshi _aža_, etc. (NC 549)
1.16. Basque _gar(h)i_ 'thin' ~ PEC _*q̇warHV_ 'narrow' (NC 933)
1.17. Basque _har,_ (R) _ãr_ 'worm' ~ PEC _*ɦabarV_ 'worm' (NC 508)
1.18. Basque _gau_ 'night' ~ Proto-Lezgian _*χ̮:am:_ > Agul _χ̮aw_-aq̰ 'evening', etc.[8]

---

[8] Lezgian forms from the Starling Caucasian Database (companion disk to NCED). _*χ̮:_ represents a tense voiceless pharyngealized uvular fricative (NCED).

1.19. Basque _ar(h)e_ ‘harrow’ ~ PEC *_ƀarhV_ = *_γ̣arhV_ ‘harrow’ (NC 477)
1.20. Basque _haga_ ‘long pole’ ~ PEC *_hăḳwV_ > Karata _hak’ʷa_ ‘branch’, etc. (NC 485)
1.21. Basque _har_-tu ‘to take, receive’ ~ Archi _kar_- ‘to take with’, etc. < PNC *_=ikĂ̄r_ (NC 632)

In several cases Basque **/a/** corresponds to Caucasian *_e_, in the environment of a liquid or (original) lateral affricate. Note some convergent forms (*_e_ > _a_) in Caucasian languages:

1.22. Basque _belarri (be-larri)_ ‘ear’ ~ Proto-Nakh *_lari_ ‘ear’, etc. < PNC *_łĕHłe_ ‘ear’ (NC 756)
1.23. Basque (Z) _k(h)arats_ ‘bitter’ ~ Archi _q’ala_ ‘bitter’, etc. < PNC *_q̇ĕɦlV_ (NC 912)
1.24. Basque _zahar_ ‘old’ ~ PNC *_swĕrho_ ‘old, year’ (NC 968)
1.25. Basque _harri_ ‘stone’ ~ PEC *_χHĕrχV_ ‘small stone, gravel’ (NC 1073)
1.26. Basque _larri_ ‘sadness, anguish’ ~ PNC *_Ł̱wĕrV_ ‘hard, severe, stern’ (NC 792)
1.27. Basque (Z) _lape_ ‘shelter of a shed’ ~ Chechen _laba_ ‘shed, peak of cap’, etc. < PEC *_ƛ̣ĕpV̄_ (NC 777)
1.28. Basque _laino, lanbro_ ‘fog, mist’ ~ PEC *_rĕnƛ̱wV_ ‘cloud, fog’ (NC 947)
1.29. Basque (L) _hardo_, (c) _arda_- ‘tinder’ ~ PEC *_ʔwēƛ̣V_ ‘a kind of grass’ (NC 230)

Basque **/e/** corresponds to Caucasian *_e_ and *_ə_:

1.30. Basque (BN,L) _bek(h)o_ ‘forehead, beak’ ~ PEC *_bĕḳwo_ ‘part of face, mouth’ (NC 289)
1.31. Basque _leka_ ‘bean pod, husk’ ~ PNC *_lĕḳV_ ‘seed, grain’ (NC 744)
1.32. Basque _bel(h)ar_ ‘grass, hay’ ~ PEC *_u̯elɣV_ ‘nettle, burdock’ (NC 1013)
1.33. Basque _erdi_ ‘half, middle’ ~ PNC *_=ĕƛ̱Ĕ̄_ ‘half, middle’ (NC 412)
1.34. Basque _el(h)e_ ‘speech, word’ ~ PEC *_lĕHwV_ ‘word’ (NC 744)
1.35. Basque _negar_ (~ _nigar_) ‘tear(s), weeping’ ~ PEC *_nĕwq̇ŭ_ ‘tear, pus’ (NC 848)
1.36. Basque _gose_ ‘hunger, hungry’ ~ PNC *_gašĕ̄_ ‘hunger’ (NC 431)
1.37. Basque _habe_ ‘pillar, beam’ ~ PEC *_hwĕbē_ ‘post, pole, tower’ (NC 497)
1.38. Basque _sare, sale_ ‘net, grate; stockade’ ~ PEC *_čɦałĕ̄_ ‘enclosure, fence’ (= **1.2**)
1.39. Basque _zelai_ ‘field, meadow’ ~ PEC *_ʒəlV_ ‘plain, plateau’ (NC 1092)
1.40. Basque (B-arc) _ze_ ‘not’[9] ~ PEC *_ʒ́ə̆_ ‘not’ (NC 1101)
1.41. Basque _lema_ ‘rudder’ ~ PEC *_ƛ̣əmʔV̆_ ‘roof’ (NC 777)
1.42. Basque _herri_ ‘inhabited place, people’ ~ PNC *_ʔwə̆hri_ ‘people, troop’ (NC 249)
1.43. Basque _behi_ ‘cow’ ~ PEC *_bħərc̱’wV_ ‘cattle’ (NC 296)
1.44. Basque _hegi_ ‘ridge’ ~ PEC *_ɦwə̄rq̱ē_ ‘mountain ridge’ (NC 536)

Basque **/e/** also corresponds to the infrequent Caucasian phoneme *_ü_:

1.45. Basque _beso_ ‘arm’ ~ PEC *_wū̆š̱V_ ‘hand, finger’ (NC 315)
1.46. Basque _abere_ / _abel_- ‘cattle’ ~ PNC *_bū̆łV_ ‘horned animal’(NC 314)

Basque **/i/** corresponds to Caucasian *_i_ and *_ɨ_:

1.47. Basque _hil_ ‘dead; die; kill’ ~ PNC *_=iwƛ̣Ĕ̄_ ‘die; kill’ (NC 661)
1.48. Basque _bizar_ ‘beard’ ~ PEC *_bilʒ́V_ ‘beard’ (NC 303)
1.49. Basque _ipini, imiñi_ ‘to put’[10] ~ PEC *_ʔima(n)_ ‘to stay, be’ (NC 210)

---

[9] Common Basque _ez_ ‘not’ (negative particle).

1.50. Basque *izeba* 'aunt' ~ PEC **=īlć̱wī* 'girl, woman' > Tsakhur *iči* 'girl', etc. (NC 952)
1.51. Basque *bi-zi* 'life, alive' ~ Lak *s:iħ* 'breath, vapor', etc. < PNC **s̱ĭHwV* (NC 961)
1.52. Basque *zikiro* 'castrated ram' ~ Karata *c':ik':er* 'kid', etc < PNC **ʒĭkV̆* 'goat, kid' (NC 1094)
1.53. Basque *hitz* 'word' ~ Chechen *=ɨc-* 'to tell', etc. < PNC **=[ī]mcŪ* (NC 642)
1.54. Basque (BN,L) *miko* 'little (bit)' ~ PEC **miḳwV* 'small, young one' (NC 821)
1.55. Basque *tipi* 'little, small' ~ Avar *hit'ína-b* 'small', etc. < PNC **ṭiHV / *HiṭV* (NC 1001)
1.56. Basque *lizun* 'moldy, musty, mustiness' ~ PEC **ƛwilc̱wV* 'dirt; bog, marsh' (NC 770)
1.57. Basque *itzuli* 'to turn, turn over, return', etc. ~ PEC **=īrcVl* 'to twirl, turn round' (NC 649)
1.58. Basque *zimitz, zimintza* 'bedbug' ~ PNC **miʒĂ / *ʒimiʒĂ* 'stinging insect' > Dargwa (Chirag) *zimizal* 'ant', etc. (NC 823)
1.59. Basque *tximitxa* [čimiča] 'bedbug' ~ PEC **ć̣imVć̣V* 'butterfly' (NC 379)
1.60. Basque *hiri* 'village, city' ~ PNC **kiħū* 'farmstead, hut' (NC 692)
1.61. Basque *ilindi, ilhinti* 'firebrand' ~ PEC **λwindV* 'firewood' (NC 764)
1.62. Basque (G) *lirdi* 'drivel, saliva' ~ PEC **λ̱wirdɨ* 'manure, pus' (NC 763)
1.63. Basque (BN,L) *k(h)ino* 'bad odor' ~ PNC **ḵwɨ̆nħV* 'smoke' (NC 738)
1.64. Basque *ihintz* 'dew' ~ Tindi *hic:u* 'bog, marsh', etc. < PEC **xwɨ̆mc̱wɨ̄* (NC 1065)
1.65. Basque (B,G,L) *lirain* 'slender' ~ PNC **=iƛ̣̃lV* 'thin' (NC 639)
1.66. Basque *ahizpa* 'sister (of a woman)' ~ PNC **=ɨ̆ć̱ɨ̆* 'sister, brother' (NC 669)
1.67. Basque *(h)osin* 'depth of water' ~ PEC **ʔwɨ̄nc̱V̆* < **ʔwic̱ɨnV* 'well, spring' (NC 232)

Basque **/o/** corresponds to Caucasian **o,* also to Caucasian **e, *ə, *i,* and **ɨ* (in labial environments). Note the convergent developments (> *o)* in some Caucasian languages:

1.68. Basque *oso* 'whole, complete' ~ PNC **=ħŏc̣V* 'full, fill' (NC 525)
1.69. Basque *atso* 'old woman' ~ PNC **c̱wŏjV* 'woman, female' (NC 374)
1.70. Basque **khola* in (BN) *gar-kola, gar-khora* 'nape' ~ PEC **qHwŏłwV* 'neck, collar' (NC 894)
1.71. Basque *ontzi* (~ *untzi*) 'vessel, container; boat, ship' ~ PEC **bōnʒ(w)V* 'vessel' (NC 311)
1.72. Basque *zor(h)i,* (B) *zoli* 'fortune, luck', etc. ~ PNC **ʒōłV* 'healthy, whole' (NC 1095)
1.73. Basque *hobi* 'gum(s) (of mouth)' ~ PNC **ħŏmG̱wɨ̆* 'throat, mouth' (NC 526)
1.74. Basque *sor-* 'body' (in compounds) ~ PEC **čōrχV* 'body' (NC 346)
1.75. Basque *otso* 'wolf' ~ PNC **bħĕrc̣ɨ̆* 'wolf' > Andi *boc'o,* etc. (NC 294)
1.76. Basque *or,* (Z) *hor, ho* 'dog' ~ PEC **χHwĕjrV-* 'dog' > Budukh *χor,* etc. (NC 1073)
1.77. Basque *erdoil, erdoi* 'rust' ~ PEC **ƛ̱wĕłʔĕ* 'mould' (NC 770)
1.78. Basque *gogor* 'hard' ~ PEC **GwērV* 'stone' (NC 467)
1.79. Basque (B) *txorru* [čor̄u] 'root of hair' ~ PEC **ć̣ħwə̄rV* 'hair' > Avar *č'or,* etc. (NC 378)
1.80. Basque (G) *alon-tza* 'mixture of grain' ~ PEC **λwɨ̆nʔɨ̄* 'seed' > Avar *xon* (NC 1021)
1.81. Basque *txori* [čori] 'bird' ~ PEC **č̣HwīlV* > Chamalal *č'or* 'bird', etc. (NC 388)
1.82. Basque *aho* 'mouth' (< **a-x*^{w}*o*) ~ PNC **χwɨ-m(V)ṗV* 'mouthful' > Khinalug *χob* (NC 1082)
1.83. Basque *ol(h)o* 'oats' ~ PNC **λwɨ̆wV* 'millet' > Chechen *hoʔ*, etc. (NC 763)

Basque **/u/** (Zuberoan **/ü/**) corresponds to Caucasian **u, *wV, *Vw.* Note the convergent developments (> *u)* in some Caucasian languages:

---

[10] Standard *ipini,* (AN, B, G) *ipiñi,* (B, BN, Sal) *imiñi* 'to put, place'. Cf. Burushaski *man-* 'to be, become', etc.

1.84. Basque *zumar* 'elm' ~ PEC **ǯɦŭmV* 'bush, quince', etc. (NC 1107)

1.85. Basque *zu* 'you' (polite) ~ PNC **źwĕ* 'you' (plural) > Lak *zu,* etc. (NC 1086)

1.86. Basque (Z) *mulho* 'petite colline, eminence, butte' ~ PEC **muɦalV* 'mountain' > Archi *mul,* etc. (NC 834)

1.87. Basque (BN) *punpula* 'tear', etc. ~ PNC **pHulq̇ɨ̆* 'dirt; secretion in the eye' (NC 871)

1.88. Basque (B,R) *un,* (L) *fuiñ* 'marrow, pith', (Z) *hün* 'brain, marrow' ~ PEC **hwĕʔnV* 'blood' (Avar 'meat') > Lak *u̯,* etc. (NC 496)[11]

1.89. Basque *zur,* (B) *zul* 'wood, timber, lumber' ~ PEC **ʒ́w[ĕ]ɫ̄* 'twig, rod, sheaf' > Avar *žul* 'broom, besom', etc. (NC 1103)

1.90. Basque *(h)ezur* 'bone' ~ PEC **rīmswe (*mswīre)* 'rib > side' > Agul *sur,* etc. (NC 954)

1.91. Basque *ahuña* 'kid' ~ PEC **kwɨ̄ʔnɨ̄* 'ram' > Andi *kun,* etc. (NC 710)

1.92. Basque (Z) *hur,* (c) *ur* 'water' ~ PEC **ɦwɨ̆lV* 'river, reservoir' > Lezgi *hül,* etc. (NC 537)

1.93. Basque (B) *zuzun* 'poplar, aspen' ~ PNC **swɨ̆nē* > Lak *sunū,* 'pomegranate' (NC 971)

1.94. Basque *guri(n)* 'butter, fat' ~ PEC **χwɨ̆lʔi* 'fat, meat' > Tabasaran *χ̣ul,* etc. (NC 1081)

1.95. Basque (AN) *zurzulo* 'nape' ~ PEC **ćwɨ̆rs̱V* 'gullet' > Dargwa (Akushi) *surs* (NC 337)

1.96. Basque *gune* 'place, space' ~ PNC **G̱winʔV* 'village, house' > Tsez *qun,* etc. (NC 471)

1.97. Basque *tuntun* 'Basque drum' ~ PNC **dwə̆nʔV* 'drum' (NC 406)

1.98. Basque *luia* 'adverse wind'[12] ~ PNC **λwəɫʔV* 'wind, breeze' > Lezgi *ful,* etc. (NC 762)

1.99. Basque *ahur* 'hollow of hand, palm' ~ PEC **kHwərV* 'pit' > Dargwa *kur,* etc. (NC 691)

1.100. Basque *zulo* 'hole, burrow' ~ PEC **ś̱wōɫV*> Avar *šulu* 'pipe', etc. (NC 978)

1.101. Basque *tutu* 'tube, pipe', etc. ~ PEC **dɦwōdwō* 'tube, pipe' > Lak *du̯du,* etc. (NC 400)

1.102. Basque *bul(h)ar* 'chest, breast' ~ PEC **Gwălɦē* 'udder, breast' (NC 465)

1.103. Basque *idulki* 'block of wood' ~ PEC **dwāɫ̄* 'stick' > Dargwa *t:ult:a* 'tree' (NC 405)

1.104. Basque *el(h)ur* 'snow' ~ PEC **λɨ̆wV* 'snow' (NC 684)

2. Lateral Affricates

It is clear that lateral affricates existed in Proto-Dene-Caucasian (PDC), and indeed are some of the most characteristic sounds of Dene-Caucasian. They are definitely reconstructed for Proto-Caucasian (where some languages retain them to the present) and for Proto-Na-Dene (where almost all languages retain them).[13] The original pattern, which is still found in many or most Na-Dene languages,[14] was a contrast of voiceless or fortis **ƛ* **[tɬ]** with glottalized **ƛ'* **[tɬ']** and voiced or lenis **λ* **[dl]**. In Basque the reflexes of all three fall together, though patterned in an interesting way:

In initial position all PDC lateral affricates **ƛ,*ƛ',*λ* > Basque **/l/**
In medial position all PDC lateral affricates **ƛ,*ƛ',*λ* > Basque **/rd/ [r̄ð]**
In final position all PDC lateral affricates **ƛ,*ƛ',*λ* > Basque **/l/**

---

[11] Basque words of the type *muin* 'marrow, brain' (*q.v.*) are of separate origin (PEC **mã̆hnū* 'brain, head'), but in Basque have been associated and contaminated with this word. The semantic series 'blood ~ flesh ~ marrow ~ brain ~ kernel' is well documented.

[12] A rather obscure word used by the Basque writer Oihenart.

[13] Peiros & Starostin (1996) also reconstruct a lateral affricate **ƛ* for Proto-Sino-Tibetan.

[14] In some Caucasian (mainly Avar-Andi-Tsezian) languages lateral affricates are retained, but not in the original *ƛ ~*ƛ' ~*λ pattern. Akhwakh, for example, has a fourfold contrast based on the features ± tense [:] and ± glottalized [']: [ƛ ~ ƛ: ~ ƛ' ~ ƛ':].

This pattern is structurally similar to that of Basque reflexes of PDC *ɬ, *q.v.*, where the contrast between **/l/** and **/lh/** only occurs between vowels. It is also structurally similar to the pattern of reflexes in Burushaski:[15]

In initial position all PDC lateral affricates *ƛ,*ƛ',*λ > Burushaski **/t/, /d/**
In medial position all PDC lateral affricates *ƛ,*ƛ',*λ > Burushaski **/lt/, /ld/**
In final position all PDC lateral affricates *ƛ,*ƛ',*λ > Burushaski **/l/**

PDC *ƛ,*ƛ',*λ > Basque initial **/l/:**

2.1. Basque *lizun* (c) 'moldy, musty, mustiness', (AN,B,G) 'dirty, untidy' ~ PEC **ƛwilc̱wV* 'dirt; bog, marsh' > Andi *ɬenc':u* 'bog, marsh', etc. (NC 770)

2.2. Basque *limuri* (Z) 'humid', (Z, Sal) 'slippery', (AN, G) 'soft, smooth', etc. ~ PEC **ƛ̱HwemV* 'liquid' > Avar *ƛ:amí*-ja-b 'liquid', Dargwa (Chirag) *šə̰m*-ze 'wet, liquid', etc. (NC 768)

2.3. Basque (B, Sal, G) *laino, laiño,* (Z) *lãnhű,* (AN-Baztan, BN-Aldude, G, L) *lanbro* 'fog, mist'[16] ~ PEC **rĕnƛ̱wV̆* 'cloud, fog' > Chechen *doχk* 'fog', Khinalug *unk'* 'cloud', etc. (NC 947) ~ Burushaski *harált* 'rain, cloud'

2.4. Basque (B,G) *loki,* (R) *lokun, lokune* 'temple (of head)' ~ PEC **ƛ̣arq̇wĕ* > Khwarshi *ƛ'oq'o* 'forehead', Avar *t'aγ̣úr* 'cap', etc. (NC 775)

2.5. Basque *lasto* 'straw (of wheat, etc.)' ~ PEC **ƛ̣ačă* 'leaf; a kind of plant' > Akhwakh *ƛ'ača* 'a kind of edible plant', etc. (NC 773)

2.6. Basque (B, G, L) *lirain* 'slender, svelte, lithe' ('esbelto [de personas y animales]') ~ PNC **=iƛ̣ĭlV* 'thin' > Avar *t'eréna*-b, Khinalug *k'ɨr* 'thin', etc. (NC 639) ~ Burushaski (H, N) *tharénum* 'narrow, tight (of clothes)'

2.7. Basque (c) *lerro* 'line, file, row' ~ PEC **ƛ̱wăr(ɦ)ə̆* 'boundary' > Avar *ƛ':er* 'garden bed, terrace, row, rank', Dargwa *jara* 'furrow', etc. (NC 782)

2.8. Basque (Z) *lape* 'shelter under the eaves of a shed' ~ PEC **ƛ̣ĕpV̄* 'stone plate or shed' > Chechen *laba* 'shed, peak of cap', etc. ~ Burushaski (H, N) *tápi* '(smaller) stone terrace' (**= 1.27**)

2.9. Basque (L) *laz* 'beam, rafter' ~ PEC **ƛ̱Vć̣V* 'log, pole' > Akhwakh *ƛ':eč'a* 'log', etc. (NC 781)

2.10. Basque (AN, B, G, L) *lema* 'rudder' (< *board, plank') ~ PEC **ƛ̣əmʔV̆* 'roof' > Karata *ƛ'ame* 'roof', etc. (NC 777)

2.11. Basque (BN, L) *lahar,* (AN, B, G) *laar,* (AN, G) *lar,* (Z) *nahar,* (Sal) *naar,* (R) *nar* 'bramble, creeping plant' ~ PNC **ƛ̣wĭrʔV* 'leaf' > Andi *ƛ'oli*, Dargwa *k'a* 'leaf', etc. (NC 784)

2.12. Basque (c) *larru,* (B) *narru* 'skin, hide, leather' ~ Dargwa (Akushi) *guli* 'skin, sheepskin', Avar *ƛ':er* 'color' (< '*skin') < PNC **Ɫ̱ŏli* (NC 789) ~ Burushaski *tar*-íŋ 'skin bag'

2.13. Basque (B) *larri* 'sadness, anxiety, anguish; sad, anxious, worried', etc. ~ Chechen *lüra* 'severe, dangerous', Bezhta *=iƛaro* 'hard', etc. < PNC **Ɫ̱wĕrV* 'hard, severe, stern' (NC 792)

---

[15] See Bengtson 1997a.

[16] Assuming metathesis: Basque *lanbro* matches well with a metathesized PEC **ƛ̱ĕnwrV̆*. It is not entirely clear that *laino, lãnhű,* and *lanbro* are one and the same word. Aulestia & White (1992) have separate entries for *laino* 'fog, mist' and *lanbro* 'dense fog'. If the Burushaski word belongs here, it may be an old compound (? **ha-* + **ralt*), or influenced by *hará-* 'to piss'.

2.14. Basque (c) *lehia, leia* 'want, wish, desire, eagerness, zeal, haste' ~ PEC **=inŁ̠wV* 'to love, want' > Chechen *laʔ-* 'to want, wish', Avar *=oƛ':-* 'to love, want, wish', Archi *ƛ':an* 'to love, want', etc. (NC 644)

PDC *ƛ,*ƛ',*λ > Basque medial **/rd/ [r̄ð]:**

2.15. Basque (B) *erdu* 'come!, come ye!' ~ PNC **=ə̠̄rƛŬ* 'to go, walk, enter' > Archi a=*ƛi-* 'to come', Lak *ulu* 'let's go!', etc. (NC 422)

2.16. Basque *erdoil, erdoi* 'rust' (of plants: AN, G), 'rust' (of iron: AN,BN,L,R,Z) ~ PEC **ƛ̠wĕłʔĕ* 'mould' > Akhwakh *xali* 'mould', Lezgi *xʷel* 'boil, furuncle', etc. (NC 770)[17]

2.17. Basque (c) *adar (a-dar < *a-r̄ðar̄)* 'horn'[18] ~ Avar *ƛ:ar,* Chechen *kur* 'horn', etc. < PEC **ƛ̠wī̆rV* (NC 771) ~ Burushaski (H) *-ltúr,* (Y) *tur* 'horn'

2.18. Basque (c) *erdi* 'half, middle' ~ PNC **=ĕƛ̠Ĕ* 'half, middle' > Bezhta *=aƛo* 'middle', *=aƛo-*kos 'half', Lak *=ač'i* 'half', etc. (NC 412)

2.19. Basque (c) *ardai, arda-,* (L) *hardo,* (Z) *ardai,* (Alava) *erdai*, etc. 'tinder (made from a kind of dried fungus)' ~ PEC **ʔwēƛ̣V* 'a kind of grass' > Chechen *jol* 'hay', Lezgi *weq'* 'grass', etc. (NC 230)[19]

2.20. Basque (AN, BN, L) *herde, erde,* (B, BN, L, R, Z) *helder, elder* 'drivel, drool' ~ PNC **ɦăṃƛă* 'sweat' > Tindi *hanla,* Lezgi *heq'* etc. (NC 509)[20]

2.21. Basque (c) *izerdi,* (B) *izardi* 'sweat', (AN,BN,R also) 'sap' ~ PEC **ć̣ā̆ŁwV* > Chechen *c'ij* 'blood', Avar *č'ágo*-b 'alive', etc. ~ Burushaski (Y) *cel,* (H,N) *chil* 'water; sap (of trees)', etc. (NC 376)

2.22. Basque *urdail* (AN, B, G, L) 'stomach', (B) 'abomasum', (B, G) 'womb, uterus'[21] ~ PEC **=īraŁ̠ V* 'stomach, abomasum, rennet' > Tindi b-*eƛ':u,* Archi b-*aƛ,* etc. (NC 670) ~ Burushaski -*úl* 'abdomen, bowels'

2.23. Basque *erdera, erdara*[22] 'foreign (of language)' / (combinatory form) *erdel-, erdal-,* as in *erdel-dun, erdal-dun* 'non-Basque-speaker, foreigner' ~ Khwarshi *ƛar* 'guest', Chechen *lū̆la-χō* 'neighbor', etc. < PEC **Ł̠ōlV* (NC 790)

2.24. Basque *mardo* (Sal, R-Uztarroz, Z) 'soft, smooth', (B, Sal, R, Z) 'robust, plump', *mardul* (B-Izpazter, G) 'vigorous, strong, exuberant / robusto, rollizo, lozano, sustancioso', *mardul*-tasun 'fertility, luxuriance' ~ PEC **mōrŁV* > Chechen *mar* 'husband', Kryz *miγil* 'male', Archi *meƛle* 'male', etc. (NC 830)

---

[17] The Basque word presupposes the development **erdoLi < *e-ƛoɨ,* or the like. **e-* is thought to be a fossilized prefix (see, *e.g.*, Bengtson 1995b.).

[18] The dissimilation of **ardar > adar* is the same as in *adore* 'courage' < Latin *ardor, ardore-* (Trask 1997, p. 145). For Basque *adar* 'branch' (a homonym) see **3.2.b.**

[19] Basque *hardo, arda-* is contaminated in some dialects with words derived from Latin *cardu(u)s* 'thistle'. For both phonetic and semantic reasons it is clear that *hardo, arda-* 'tinder' and Basque words such as *kardu, gardu* 'thistle' are of entirely distinct origins.

[20] The Basque word *(h)elder* above is easily confused and contaminated with Basque *lerde, lirdi* 'drivel, saliva', *q.v.,* of a quite separate origin ~ PEC **ł̠wirdɨ* (LDC #19). Note also Basque (G) *bilder (bi-lder)* 'drivel', which apparently incorporates an old class marker *(*bi-/be-) + helder.*

[21] Caucasian has **=īraŁ̠ V,* so assimilation or dissimilation is assumed. Lak *č:arlu* 'kidney' requires metathesis (< PEC **=īŁ̠arV),* and assimilation would produce PEC **=īŁ̠alV,* close to the Basque form. The Basque word also has the fossilized class prefix *u-*.

[22] Generally, *erdera / erdel-* to the south and west of Basque country, *erdara / erdal-* to the north and east.

PDC *ƛ,*ƛ',*λ > Basque final **/l/**. At present I can cite only two secure examples:

2.25. Basque (BN, L, Z) *oihal,* (AN, B, G, R) *oial* 'cloth, fabric' (< **o(i)-xal*) ~ Bezhta *χiƛo* 'trousers, breeches', Lak *harč:ala* 'cuff, trouser leg', etc. < PEC **χwĭlȽ̱V* (NC 1081)
2.26. Basque (c) *hil, il* 'dead, death, die' ~ PNC *=*iwƛ̣Ĕ* 'to die, kill' > Chechen =*al-* 'to die', Bezhta *-iƛ'-* 'to kill', etc. (NC 661)

In the following cases we find the unusual (non-trivial) correspondence of Basque **/l/** = Burushaski **/t/ (~ /th/)**:

2.27. Basque *limuri* (Z) 'humid' ~ Burushaski *tam* dél- 'to bathe', etc. ~ PEC **ƛ̱HwemV* 'liquid' (= **2.2**)
2.28. Basque *lirain* 'slender' ~ Burushaski (H) *tharén*-um 'narrow' ~ PNC *=*iƛ̣̄lV* 'thin' (= **2.6**)
2.29. Basque (Z) *lape* 'shelter under the eaves of a shed' ~ Burushaski *tápi* 'Felsterrasse' ~ PEC **ƛ̣ĕpV̄* 'stone plate or shed' (= **2.8**)
2.30. Basque *larru* 'skin, leather' ~ Burushaski *tar*-íŋ 'skin bag' ~ PNC **Ƚ̱ŏli* 'skin, color' (= **2.12**)
2.31. Basque *leku, lekhu* 'place' ~ Burushaski *tik* 'earth, ground; rust'[23]
2.32. Basque *larre* (AN,BN,G,L,Z), *larra* (B) 'grassland, pasture', also (BN) 'heath' ~ Burushaski (Y) *ter* 'summer pasture' (Berger: 'Hochweide, auf die das Vieh im Sommer etrieben wird.')

In five of the six cases there is evidence from other Dene-Caucasian languages that the original initial consonant was a lateral affricate. In the sixth case (**2.32**) the pattern of correspondence suggests that the original initial was a lateral affricate, though I have found no direct evidence of one, since I do not know of any cognate in Caucasian or Na-Dene.

3. Basque Homonyms

We are dealing with comparisons between a language with a relatively simple phonetic system (Basque), and languages that are much more complex (Caucasian: especially as regards consonants). As seen at the beginning of this paper, Basque has about 23 consonant phonemes, while Proto-Caucasian had about twice as many. It is thus to be expected that some words that have phonetically merged in Basque would correspond to words that are distinct in Proto-Caucasian, and this is indeed the case, according to my findings. The examples discussed here are Basque *bel(h)ar* 'forehead' : *bel(h)ar* 'grass, hay'; *adar* 'horn' : *adar* 'branch'; *eho* 'grind' : *eho* 'beat' : *eho* 'weave'; *hobi* 'gum(s)' : *hobi* 'grave'. The distinct origins of these homonyms are shown by the comparisons below:

3.1.a. Basque (BN, Sal, Z) *belhar* 'forehead' (< **belha-r)* ~ Rutul *bäl,* Lezgi *p:el,* Kryz *bel,* Budukh *beliǯ* 'forehead', Archi *bat* 'horn', Tindi *bala* 'edge, end, corner', etc. < PEC **bʕāɬhŏ* (NC 285)[24] ~ Burushaski *bal* 'wall'(< '*edge')
3.1.b. Basque (BN, L, Z) *belhar* '(first mowing of) hay', (AN, G, L) *belar* 'grass, hay', (B, G) *berar,* (B) *bedar* 'grass, hay' ~ Lezgi *werg* 'nettle', Archi *urk:i* 'burdock', etc. < PEC **ṷelɣV* (NC 1013)

---

[23] Cf. also: PST **ƛ[ia]k* 'iron' > Tibetan *lčag*-s 'iron', etc. (ST III:68); Na-Dene: Haida (A) *ƛak* 'land, place', (S) *ƛ̥́ga* 'land'; Tlingit *ƛ'έχ'kw* ~ *ƛ'áχ'kw* 'soil'; Sarsi gù-*ƛ'ìs* 'earth').
[24] NCED proposes the reconstructed meaning 'edge, end', whence 'forehead, horn' in some languages.

3.2.a. Basque *adar (a-dar < *a-r̄ðar̄)* 'horn'[25] ~ Avar *ƛ:ar,* Chechen *kur* 'horn', etc. < PEC **ƛ̱wɨr̆V* (NC 771) ~ Burushaski (H) *-ltúr,* (Y) *tur* 'horn'
3.2.b. Basque *adar* 'branch' (< **ar̄ðar̄*)[26] ~ Avar *ʕarƛ̄':él* 'branch, bough', Tsez *aƛ̄'iru* 'pod', etc. < PEC **ɦălƛ̱VɫV* (NC 508) ~ Burushaski (H, N) *yáltar* 'the upper leafy branches of a tree'

3.3.a. Basque (c) *eho,* (B) *eio* 'to grind', (c) *eihera* 'mill', etc. ~ Ingush *aħ-* 'to grind', *ħajra* 'mill', Andi *ʔiχʷo*-qi- 'to grind', etc. < PNC **Hĕmχ̱wV* 'to grind' (NC 559)
3.3.b. Basque (c) *eho* 'to beat', (Z) 'to kill; put lights out', (AN) *eo* 'give (someone) a thrashing, beat (someone) soundly', (B) *eio* 'to get very tired, get fatigued', etc. ~ Avar *=uχ-* 'to beat, hit', Rutul *=ɨχ̯a-* 'to beat, hit', etc. < PNC **HīrχA* (NC 581)
3.3.c. Basque (c) *eho,* 'to weave', *ehun* 'to weave; cloth', (B) *eio* 'to braid' ~ PEC **=irχ̱wVn* 'to knit, weave, spin' > Chamalal *χ:in-* 'to spin', Dargwa *=imχ-/=umχ-* 'to plait, weave', Agul *ruχ-* 'to knit', etc. (NC 655)

3.4.a. Basque (BN, L, Z) *hobi,* (B, G) *oi* 'gum(s) (of mouth)' ~ Akhwakh *oq':o* 'throat', Tsez *haqu* '(inside of) mouth', Circassian *ʔʷə* 'mouth', etc. < PNC **ɦŏmG̱wĭ* (NC 526)
3.4.b. Basque (c) *hobi* 'grave, tomb' ~ PEC **fïwɨ* 'grave' > Avar *χob,* Tindi *hoba,* Lak *haw,* etc. (NC 428)

In the case of **3.2.a.** and **3.2.b.** both words are preserved in all three Macro-Caucasian branches, but only in Basque do the words merge phonetically

[**3.2.a.** Basque *adar* 'horn'	: Avar *ƛ̄:ar* 'horn'	: Burushaski *-ltúr* 'horn']
[**3.2.b.** Basque *adar* 'branch'	: Avar *ʕarƛ̄':él* 'branch'	: Burushaski *yáltar* 'branches']

Conclusions

Luis Michelena (1961, *et al.*) provided us a tremendous service in cataloguing the diverse phonetic forms of the Basque dialects, and giving us some direction in understanding the changes. Michelena did not accept the relationship between Basque and Caucasian, and thus could not apply external comparison to the problems of Basque phonology. In Michelena's defense I would point out that in his lifetime the Basque-Caucasian comparisons that had been published were very uneven in quality and haphazard, and even the available materials on Caucasian languages were very sparse, with no deep reconstructions. We now have a significant advantage because of new sources and reconstructions of Caucasian languages that are now available (*e.g.,* Nikolayev & Starostin 1994, Chirikba 1996). For Burushaski we now have the definitive books by Berger (1974, 1998). With these materials we now have a solid basis of what to compare from the Caucasian and Burushaski side.

This essay is only a beginning. The extensive comparisons involving consonants will be published separately.[27] (Only the reflexes of PDC lateral affricates were discussed in depth here.) However, I suggest that the evidence presented here is already strongly indicative that the proposed genetic relationship between Basque and Caucasian is highly probable, and that it can be verified by numerous recurrent phonological correspondences.[28]

---

[25] The dissimilation of **ardar > adar* is analogous to that in Basque *adore* 'courage' < Latin *ardor, ardore-* (Trask 1997, p. 145).
[26] There is also a Basque variant with **/b/,** *abar* 'branch', which is difficult to explain.
[27] These would add about 60 pages to this paper.

Abbreviations

AN	Alto Navarro = *Nafarroa Garaia* = High Navarrese (Basque dialect)
arc	archaic
B	*Bizkaia* = Biscayan (Basque dialect)
BN	Basse Navarre = *Nafarroa Beherea* = Low Navarrese (Basque dialect)
c	common (Basque) = *Euskara Batua*
CWC	*Common West Caucasian* = Chirikba 1996
DC	Dene-Caucasian = Sino-Caucasian
G	*Gipuzkoa* = Guipúzcoa (Basque dialect)
H	Hunza (Burushaski dialect)
L	*Lapurdi* = Labourdin (Basque dialect)
LDC	"Lexica Dene-Caucasica" = Blažek & Bengtson 1995
N	Nagiri, Nagari (Burushaski dialect); northern (Basque)
NC	North Caucasian; in reference citations = NCED
NCED	*North Caucasian Etymological Dictionary* = Nikolayev & Starostin 1994
PA	Proto-Athabaskan
PDC	Proto-Dene-Caucasian
PEA	Proto-Eyak-Athabaskan
PEC	Proto-East Caucasian (= Proto-Nakh-Daghestanian)
PNC	Proto-(North) Caucasian
PND	Proto-Na-Dene
PST	Proto-Sino-Tibetan
PWC	Proto-West Caucasian (= Proto-Abkhazo-Adygean)
PY	Proto-Yeniseian
R	Roncalese = *Erronkari* (Basque dialect)
S	southern (Basque)
Sal	Salazarese = *Saraitza* (Basque dialect)
SSEJ	"Sravnitel'nyj slovar' enisejskix jazykov" = Starostin 1995
ST	*(A Comparative Vocabulary of Five) Sino-Tibetan (Languages)* = Peiros & Starostin 1996
Y	Yasin(i) = Werchikwar (Burushaski dialect)
Z	*Zuberoa* = Souletin (Basque dialect)

References

Aulestia, Gorka, and Linda White, eds. 1992. *Basque-English English-Basque Dictionary.* Reno, Las Vegas, London: University of Nevada Press.

Azkue, R.M. de. 1905. *Diccionario vasco-español-francés.* Bilbao.

Bengtson, John D. 1994. "On the Genetic Classification of Basque." *Mother Tongue* (Newsletter) 22: 31-36.

---------- 1995a. "Is Basque Isolated? (ad Trask [1994-95])." *Dhumbadji!* 2.2: 33-44.

---------- 1995b. "Basque: An orphan forever?" *Mother Tongue* (Journal) 1: 84-103.

---------- 1996. "Correspondences of Basque and Caucasic Final Vowels: -i/-e, -u/-o." *Fontes Linguae Vasconum* 71: 7-15.

---------- 1997a. "Ein Vergleich von Buruschaski und Nordkaukasisch." *Georgica* 20: 88-94.

---------- 1997b. "Basque and the Other Dene-Caucasic Languages." In *The Twenty-Third*

---

[28] See the index of Basque words (Appendix B), in which many words exhibit more than one correspondence with Caucasian.

LACUS Forum, ed. by A.K. Melby, pp. 137-148. Chapel Hill, NC: LACUS.
---------- 1998a. "Dene-Caucasian 'Navel': some proposed etymologies." *Dhumbadji!* 4.1: 86-90.
---------- 1998b. "Some Yeniseian Isoglosses." *Mother Tongue* (Journal) 4: 27-32.
---------- 1999a. "A Comparison of Basque and (North) Caucasian Basic Vocabulary." *Mother Tongue* (Journal) 5: 40-57.
---------- 1999b. "'Eye, Ear, Tongue' in Basque and East Caucasian." In *From Neanderthal to Easter Island (Tribute to W.W. Schuhmacher),* ed. by N.A. Kirk and P.J. Sidwell, pp. 3-10. Melbourne: Association for the History of Language.
---------- 1999c. Review of Trask 1997. *Romance Philology* 52: 219-224.
Berger, Hermann. 1974. *Das Yasin-Burushaski (Werchikwar).* Wiesbaden: Harrassowitz.
---------- 1998. *Die Burushaski-Sprache von Hunza und Nager.* 3 vols. Wiesbaden: Harrassowitz.
Blažek, Václav. 1995. "Towards the Position of Basque: a reply to Trask's critique of the Dene-Caucasian hypothesis." *Mother Tongue* (Journal) 1: 104-110.
Blažek, Václav, and John D. Bengtson. 1995. "Lexica Dene-Caucasica." *Central Asiatic Journal* 39.1: 11-50. (Additions and corrections in *CAJ* 39.2: 161-164.)
Catford, J.C. 1977. "Mountain of Tongues: The Languages of the Caucasus." *Annual Review of Anthropology* 6: 283-314.
Chirikba, Viacheslav A. 1996. *Common West Caucasian.* Leiden: Research School CNWS.
Diakonoff, Igor M., and Sergei A. Starostin. 1986. *Hurro-Urartian as an Eastern Caucasian Language.* Munich: R. Kitzinger.
Hualde, José Ignacio. 1991. *Basque Phonology.* London/New York: Routledge.
Kuipers, A.H. 1975. *A Dictionary of Proto-Circassian Roots.* Lisse, Netherlands.
Larrasquet, J. 1939. *Le Basque de la Basse-Soule orientale.* Paris.
Michelena, Luis. 1961. *Fonética Histórica Vasca.* San Sebastián: Diputación de Guipúzcoa.
Moutard, Nicole. 1975. "Étude phonologique sur les dialects basques, I." *Fontes Linguae Vasconum* 19: 5-42.
Nikolayev, S.L., and S.A. Starostin. 1991. "North Caucasian Roots." In Shevoroshkin 1991: 174 -264.
---------- 1994. *A North Caucasian Etymological Dictionary.* Moscow: Asterisk Press.
Shevoroshkin, Vitaly, ed. 1991. *Dene-Sino-Caucasian Languages.* Bochum: Brockmeyer.
Starostin, S[ergei] A. 1982. Праенисейская реконструкция и внешние связи енисейских языков. [Proto-Yeniseian reconstruction and external relations of the Yeniseian languages.] In *Кетский Сборник (Studia Ketica),* pp. 144-237. Leningrad: Nauka.
---------- 1984. Гипотеза о генетических связях синотибетских языков с енисейскими и севернокавказскими языками. In *Лингвистическая реконструкция и древнейшая история Востока,* часть 4, pp. 19-38. Moscow: Akademija Nauk, Institut Vostokovedenija. [see Starostin 1991]
---------- 1991. "On the Hypothesis of a Genetic Connection Between the Sino-Tibetan Languages and the Yeniseian and North Caucasian Languages." In Shevoroshkin 1991: 12-41. [Translation of Starostin 1984]
---------- 1995. Сравнительный словарь енисейских языков. [Comparative dictionary of the Yeniseian languages.] In *Кетский Сборник (Studia Ketica),* ed. by S.A. Starostin, pp. 176-315. Moscow: Shkola Jazyki Russkoj Kul'tury. {SSEJ}
---------- 1996. "Comments on the Basque-Dene-Caucasian Comparisons." *Mother Tongue* (Journal) 2: 101-109.
---------- 2002. "North Caucasian Morphology." (ms.)

Trask, R.L. 1994-95. "Basque: The Search for Relatives (Part 1)." *Dhumbadji!* 2.1: 3-54.
---------- 1995. "Basque and Dene-Caucasian: A Critique from the Basque Side." *Mother Tongue* (Journal) 1: 3-82.
---------- 1997. *The History of Basque.* London/New York: Routledge.

Appendix A:
Key to Phonetic Characters

ã ẽ ĩ õ ũ	nasal vowels
c: k: ƛ: p: s: t:, etc.	tense obstruents
c' k' ƛ' p' t', etc.	glottalized obstruents (NCED c̣ ḳ ƛ̣ ṗ ṭ , etc.)
ć ć' ʒ́	hissing-hushing affricates (NCED ć ć̣ ʒ́)
č č' ǯ	hushing affricates (NCED č č̣ ǯ)
ð	voiced dental fricative: e.g., Basque *adar* [aðar̄] 'horn', *erdi* [er̄ði] 'half, middle'
ɢ	voiced uvular stop
γ	voiced velar fricative (Caucasian, Na-Dene); voiced uvular fricative (Burushaski)
γ̣	voiced uvular fricative (NCED ʁ́)
ħ	voiceless emphatic laryngeal fricative
ɦ	voiced laryngeal fricative
ɫ	lateral resonant or glide (in PNC and PST reconstructions)
ɬ	voiceless lateral fricative (NCED λ)
L	voiced lateral fricative in NCED (= ɮ)
Ł	voiced lateral affricate (NCED)
λ	voiceless lateral fricative in NCED (= ɬ)
ƛ	voiceless lateral affricate
ƛ́	palatalized voiceless lateral affricate
ƛ'	glottalized voiceless lateral affricate (NCED ƛ̣)
ƛ́'	palatalized glottalized lateral affricate
ʎ	palatalized lateral resonant
ɮ	voiced lateral fricative (NCED L)
ŕ	palatalized rhotic resonant
r̄	rhotic trill (= Basque *rr*)
ʁ	voiced uvular fricative (= Burushaski *γ*)
ś	voiceless hissing-hushing fricative
š	voiceless hushing fricative
ṣ	voiceless retroflex fricative (Burushaski)
ṭ	voiceless retroflex stop (Burushaski)
u̯	labial glide (in Caucasian reconstructions)
x	voiceless uvular fricative (Burushaski); voiceless velar fricative (Caucasian, Na-Dene)

χ	voiceless uvular fricative (= Burushaski *x*)
χ̮	pharyngealized voiceless uvular fricative (NCED χI)
ẏ	retroflex velarized spirant (Burushaski)
ʒ	voiced hissing affricate (= d͜z)
ʒ́	voiced hissing-hushing affricate (= d͜ź)
ǯ	voiced hushing affricate (= d͜ž) (= Burushaski *j*)
ʔ	glottal stop
ʡ	glottalized emphatic laryngeal stop (Caucasian)
ʕ	voiced emphatic laryngeal fricative

Appendix B:
Index of Basque Words Cited

Words are generally cited in their common Basque (*euskara batua*) form, following Aulestia & White (1992). Some dialectal forms, especially Bizkaian and Zuberoan, are also cited with a cross-reference to the common Basque form (*e.g., aizta* see *ahizpa*). Phonemic **/h/** that is pronounced in northern Basque dialects but not found in standard Basque spelling is shown as *(h)*, *e.g., k(h)arats* 'bitter', *ol(h)o* 'oats'.

a-, -a *see* ha-0
abere 'domestic animal(s), cattle' **1.46**
adar 'branch' **3.2.b**
adar 'horn' **2.17, 3.2.a**
ahizpa 'sister (of a woman)' **1.66**
aho 'mouth' **1.82**
ahuña, ahuñe 'kid' **1.91**
ahur 'palm, hollow (of hand)' **1.99**
aizta *see* ahizpa
al(h)o *see* ol(h)o
alontza 'mixture of grain' **1.80**
apal 'shelf' **1.11**
apo 'hoof' **1.4**
ardai 'tinder' **1.29, 2.19**
ar(h)e 'harrow, rake' **1.19**
atso 'old woman' **1.69**
atze 'tree' **1.15**
aundi *see* handi

beharri *see* belarri
behatz 'toe, thumb' **1.3**
behi 'cow' **1.43**
bek(h)o 'forehead' **1.30**
belarri 'ear' **1.22**
bel(h)ar 'forehead' **3.1.a**
bel(h)ar 'grass, hay' **1.32, 3.1.b**
beso 'arm' **1.45**
bilder *see* elder
bizar 'beard' **1.48**
bizi 'to live; alive; life' **1.51**
budar *see* bul(h)ar
bul(h)ar 'breast, chest' **1.102**

edur *see* el(h)ur
eho ‘to grind’ **3.3.a**
eho ‘to beat’ **3.3.b**
eho, ehun ‘to weave’ **3.3.c**
eihera ‘mill’ **3.3.a**
eio *see* eho
elder, erde ‘drool, drivel’ **2.20**
el(h)e ‘speech, story, word’ **1.34**
el(h)ur ‘snow’ **1.104**
erde *see* elder
erdera, erdara ‘foreign language’ **2.23**
erdi ‘half, middle’ **1.33, 2.18**
erdoil ‘rust’ **1.77, 2.16**
erdu ‘come ye!’ **2.15**
ez ‘not’ **1.40**

fuiñ *see* (h)un, hün

gar(h)i ‘thin’ **1.16**
garkola, garkhora ‘nape’ **1.70**
gau ‘night’ **1.18**
gogor ‘hard’ **1.78**
gor(h)i *see* gurin
gose ‘hunger, hungry’ **1.36**
gune ‘place, space’ **1.96**
gurin, gur(h)i, gor(h)i ‘butter, fat’ **1.94**

ha-, -a article, demonstrative **1.1**
habe ‘pillar, beam’ **1.37**
haga ‘long pole, rod’ **1.20**
handi ‘big, great’ **1.13**
har ‘worm’ **1.17**
hardo *see* ardai
hari ‘thread, string, wire’ **1.12**
harri ‘stone, rock’ **1.25**
har-tu ‘to take, receive’ **1..21**
hatz ‘finger’ **1.3**
hegi ‘edge, ridge’ **1.44**
helder, herde *see* elder, erde
herri ‘town, people, nation’ **1.42**
hezur ‘bone’ **1.90**
hil ‘die; dead; death’ **1.47, 2.26**
hiri ‘village, city’ **1.60**
hitz ‘word’ **1.53**
hobi ‘grave’ **3.4.b**
hobi ‘gum(s)’ **1.73, 3.4.a**
hor, ho *see* or
hosin *see* osin
(h)un, hün ‘marrow, pith, brain’ **1.88**
hur *see* ur

idulki ‘block of wood’ **1.103**
ihintz ‘dew’ **1.64**
il(h)inti, ilindi ‘firebrand’ **1.61**
imiñi *see* ipini

ipini 'to put, place' **1.49**
itzuli 'to turn' **1.57**
izar 'star' **1.6**
izardi *see* izerdi
izeba 'aunt' **1.50**
izerdi 'sweat, sap' **2.21**

k(h)arats 'bitter, sour' **1.23**
k(h)ino 'bad odor, bad taste' **1.63**

lahar 'bramble' **2.11**
laino 'fog, mist' **1.28, 2.3**
lamika-tu 'to lick' **1.9**
lanbro *see* laino
lape 'shelter' **1.27, 2.8, 2.29**
larre, larra 'grassland, pasture' **2.32**
larri 'worried, serious', etc. **1.26, 2.13**
larru 'skin, leather' **2.12, 2.30**
lasto 'straw, hay' **1.8, 2.5**
laz 'beam, rafter' **2.9**
lehia 'diligence, laboriousness', etc. **2.14**
leka 'bean pod, husk' **1.31**
lek(h)u 'place' **2.31**
lema 'rudder' **1.41, 2.10**
lerde 'drivel, drool' **1.62**
lerro 'line, file, row' **2.7**
limuri 'slippery; humid' **2.2, 2.27**
lirain 'slender, svelte' **1.65, 2.6, 2.28**
lirdi *see* lerde
lizun 'mold, mildew' **1.56, 2.1**
loki, lokun 'temple (of head)' **2.4**
luia 'adverse wind' **1.98**

mardo 'luxuriant, vigorous; soft, smooth' **2.24**
mardul 'robust, healthy, strong' **2.24**
miko 'a little bit' **1.54**
muin, muñ 'pith, marrow, brain' cf. **1.88** (footnote)
mul(h)o 'heap, mound' **1.86**

negar, nigar 'weeping, tears' **1.35**

oi *see* ohe, hobi
oihal 'cloth' **2.25**
ol(h)o 'oats' **1.83**
ontzi 'vessel' **1.71**
or 'dog' **1.76**
osin 'well, depth' **1.67**
oso 'whole, complete' **1.68**
otso 'wolf' **1.75**

punpu(i)l(l)a 'tear, blister, bubble' **1..87**

sabel 'belly' **1.7**
sagu 'mouse' **1.10**
sare, sale 'net(-work)' **1.2, 1.38**
sor- 'body' **1.74**

sor-balda 'shoulder' **1.74**

takoin 'heel (of shoe)' **1.14**
tipi 'little, small' **1.55**
tuntun 'Basque drum' **1.97**
tutu 'tube, pipe' **1.101**

tximitxa 'bedbug' **3.53, 8.17, 9.59**
txori 'bird' **1.81**
txorru 'root of hair' **1.79**

udagara 'otter' **1.5**
ugabere, ugadara, ugadera *see* udagara
un *see* (h)un
une *see* gune
untzi *see* ontzi
ur 'water' **1.92**
urdail 'stomach, abomasum, womb' **2.22**
urde '(male) pig' **8.100**
uri *see* hiri
urin *see* gurin

xori *see* txori

zahar 'old' **1.24**
ze 'not' **1.40**
zelai 'meadow, plain' **1.39**
zik(h)iro 'castrated ram' **1.52**
zimitz, zimintza 'bedbug' **1.58**
zoli *see* zor(h)i
zor(h)i 'luck; mature' **1.72**
zu 'you' **1.85**
zulo 'hole, burrow' **1.100**
zumar 'elm' **1.84**
zur, zul 'wood, timber, lumber' **1.89**
zurzulo, zurzuil 'nape' **1.95**
zuzun 'poplar, aspen' **1.93**

Salishan and North Caucasian

Vitaly V. Shevoroshkin
University of Michigan,
ASLIP Council of Fellows

Introduction

Salishan [Sa] languages are represented in this paper in the following way (data used in this paper are taken from the appropriate dictionaries or published word lists):

Tsamosan [Ts]: Upper Chehalis [UP].

Interior Salish [IS]: Thompson River Salish [Th], Shuswap [Sh], Colville-Okanagan [CO], Moses-Columbian [MC], Spokane [Sp], Montana Salish [MS].

Central Salish [CS]: Lushootseed [Ls] (=Puget), Sechelt [Se], Squamish [Sq].

Bella Coola (Nuxalk) [BC]/[Nu].

I also occasionally use North Wakashan [NWk] language data as provided by N. Lincoln and J.C. Rath in their *North Wakashan Comparative Root List* (Ottawa 1980); abbreviations: Haisla = Ha; Heiltsuk = He; Kwakiutl (=Kwakwala) = Kw; Oowekyala = Oo. - Note also: Wakashan = Wk; Makah = Ma; Nitinat = Ni; Nootka = No.

North Caucasian [NC] languages are cited after *A North Caucasian Etymological Dictionary* by S.L. Nikolaev and S.A. Starostin (Asterisk Publishers, Moscow 1994) [NCED], and occasionally also after S.A. St[arostin]'s papers "Nostratic and Sino-Caucasian" (in *Explorations in Language Macrofamilies*, Bochum 1989: 42-66) [St. '89] and "On the Hypothesis of a Genetic Connection between the Sino-Tibetan Languages and the Yeniseian and North-Caucasian Languages" (in *Dene-Sino-Caucasian Languages*, Bochum 1991: 12-40) [St. '91]. Note relevant abbreviations: Northeast Caucasian = NEC = EC; Northwest Caucasian = NWC = WC; Sino-Caucasian = SC; Yeniseian = Yen; Sino-Tibetan = ST; Nostratic = N.

Abbreviations of NC daughter language designations follow the pattern adopted in NCED; in addition, I provide abbreviations of designations of some especially important NEC and NWC daughter languages in the text of this paper.

Sa languages show a remarkable uniformity of their sound systems,[1] making shallow the existing PSa reconstruction and "elevating" archaic languages (which have preserved retracted sounds, and did not palatalize *k k' x*) practically to the level of PSa.[2]

---

[1] Sa stops *p t c k* k^{w} *q* q^{w} have glottalized counterparts; this is also valid for *m n r l y w*. - Stops *k* k^{w} *q* q^{w} have also appropriate fricative counterparts *x* x^{w} *x̣* $x̣^{w}$; voiced: *ɣ* (velar?), *ʕ* $ʕ^{w}$ (either voiced uvulars or pharyngeals). - There are also *ɬ ƛ' h ʔ*. - On some occasions, Sa voiced consonants *z z' ɣ ʕ* $ʕ^{w}$ seem to match NC voiced consonants (see examples below).

[2] PSa had at least three vowels as well as their retracted counterparts: *a ạ i ị u ụ* [These vowels participate in an old ablaut *a/i*; *a/u*; etc.; cf. ablaut in the NC languages; Sa *i* is considered the

I. Ways to Compare Salishan (and Wakashan) Languages with North Caucasian Languages

Both Sa and Wk languages show many very strong genetic ties with NC languages (a relationship which seems much more intimate than that between NC on the one hand, and both Yen and ST on the other).[3]

It is totally inappropriate to deny mutual genetic relationship between Sa and Wk just because there exist many borrowings from Wk to Sa, and *vice versa*. There are many identical, or semi-identical, Wk-Sa word pairs that belong to the most stable lexicon (first and second person pronouns; some body part denotations, etc.); these root correspondences cannot be considered borrowings, simply because such words are not subject to borrowing. Such correspondences indicate deep genetic relationship.

This paper deals primarily with Sa - NC cognates; some Sa - Wk, Wk - NC, and Sa - Wk - NC cognate sets are also present.

Sa (and Wk) sounds either match NC sounds directly, or in a way which shows that Sa sounds are "reduced" representations of NC sound combinations (a given Sa sound may represent several sounds of a much larger sound system; original sounds may disappear; original clusters may be reduced to simple consonants, or eliminated in the following way: **CC > CVC*, or **CC > C...C*). Relatively frequent metatheses of Sa roots, as compared with NC roots, are similar to frequent metatheses of NC roots. (This may be reflected in reconstruction of 2 variants of a given root, or a given NC root may be metathesized in appropriate daughter languages).

There is no sufficient reconstruction of PSa; the existing reconstructions of NWk roots in many cases seem to be incorrect. On the other hand, both Sa and Wk languages have clearly preserved many features of the underlying sound system: cf. preservation of q q' q^{w} q'^{w}; x $\underset{\cdot}{x}$ x^{w} $\underset{\cdot}{x}^{w}$; *l l' ł λ* [Wk only] *ƛ* [Wk only] *ƛ'* (etc.). There are no significant differences between genetically related roots in various Sa (or, for that matter, Wk) languages.

Scholars agree that there were inherited voiced consonants in Sa; they seem to be best preserved in IS:Th: *z ʒ ɣ ʕ* $ʕ^{w}$ (cf. also Wk). The following examples 1-4 seem to show that Sa voiced consonants may match NC voiced counterparts in genetically related roots/words:

---

stablest vowel]. Historically, retraction indicates a back articulation in the original root. Several consonants have retracted counterparts: *s c l l'* vs *ṣ c̣ ḷ ḷ'*. Always retracted are *r r'* and *z z'*. Most Sa languages show no retraction; most languages lack consonants *r r' z z'*.

[3] When comparing languages on a broader scale, - i.e., not just Sa (or Sa-Wk) vs NC, but Sa, Wk, Ath[apascan] (etc.) vs SC (or Yen, or ST, for that matter) vs N (or Kartvelian, for that matter), - one can use SC data even in cases where there are no NC cognates, - for instance (N.Kruglyj-Enke, Moscow 2000 Conference on deep reconstruction): FIRE, BURN: Sa $*p'i\underset{\cdot}{x}(^{w})$ 'fire, burn' // SC **piHwV* 'heat' (**p < *p'*) // N **p'a/iyxwV* 'fire'. In the present paper such broad comparisons appear very seldom.

(1) STINGING INSECT: IS:Th *məz·/mə̣ze* 'flies'; *məc'·/mə̣c'e* 'bees, hornets, wasps' // NC **miʒA̋* 'stinging insect'. [Alternations of the type *c/z* are typical also for Wk].

(2) LYNX (etc.): PSa **mVɣawʔ* (Kuipers: **(s-)mɣaw(ʔ)* 'feline, coyote') > IS:Th *məɣew',* Sh *s-mɣewʔ* // NEC **mHarGVwV* 'tom-cat' (> Lezg. **marq:/aw* > Ag. **maʁ/aw* / Tsez. **maʁ:ur*, also with **-χ-*). [Cf. NWk:He *mauχ̣wa* 'bob-cat'].

(3) WORM: IS:CO *m'aʕ-mlaʔ* (metathesis + partial redupl.?) // NC **mHilaGwV.*

(4) GREASE: IS:Th *miʕ'*^{W}*-* (root) // NEC **măɦwV*; etc..

The above examples show that a relatively close genetic relationship between Sa and NC may be seen rather clearly even if we take only one Sa language and compare it to NC. [Naturally, when we deal with several Sa languages (which have preserved a given root) the comparison will look more solid].

In example 5 a PSa root is represented by several Sa languages; the NEC match is exact. Example 6 shows only one Sa language, which is archaic (actually, it is more archaic in this case than even NEC which has lost initial **t-* (NEC **q'Hwā̆*); this **t-* is still present in NWC). Both examples 6 and 7 show semi-identical matches between MC (an archaic Sa language) and NWC Example 8 provides a precise match between Sa (UC), Wk (Kw), and NEC languages:

(5) DRINK: Ts:UC *q*^{W}*o·ʔ* (from PSa; cf. Th *ʔu-q*^{W}*eʔ*, Ls *q*^{W}*uʔ*, etc.) // NEC **ʔu-qwV.*

(6) TWO: IS:MC *tq'aw'-s* (cf. *t'q'*^{W}*maw'-s-ən* 'together') // NC **tq'Hwǟ* > NWC **tq/ ':*^{W}*A* (> Ubykh *ṭq̇*^{W}*a*). [Sa may show a typical vowel-insertion: *CVC* for *CC*.]

(7) TREE: IS:MC *c'əl* // NWC **c̣əla* < NC **c̱'ə̆/ŏłV.*

(8) DIRT: Ts:UC *ciq*^{W}*-i-* // NWk *cq*^{W}*-*, *zq*^{W}*-*; *c'q*^{W}*-* 'dirty'; cf. Kw *ʒeq*^{W}*á̆* 'dirt' // NEC **č̱Hɨ̆q'wĂ.*

In Sa - NC comparison, some existing reconstructions (mostly proposed by A. Kuipers; cf. *Lingua* v. 57, 1982: 93-100) can be used, but we may note that important phonetic elements (which still appear in some remnants of underlying consonants or consonant clusters) may be absent in these reconstructions. Unfortunately, the number of the existing reconstructions is very small; on occasion, these reconstructions are not satisfactory from the point of view of comparative and diachronic semantics. (There is a very strong tendency among linguists working on Sa and, especially, Wk languages to genetically unite different, though phonetically somewhat similar, roots, however improbable such a tie-in may be from the point of view of historical linguistics and semantics.)

Nevertheless, many existing PSa reconstructions are quite acceptable; it is not by chance that these reconstructions often match NC roots in a very precise manner:

(9) NECK, THROAT, GOITRE: PSa **qənu/ax*^{W} 'throat, gullet' (Kuipers) // NEC **q'*^{W}*i/ēnwV* 'goitre, Adam's apple'. [Cf. NWk:He *q'*^{W}*ú-q'*^{W}*ùni* 'neck'. - Wk may reflect a process of simplification: **q'*^{W} > *q'/q*^{W} > *q*, cf. Sa *q* in all languages; some NEC languages show *q'*].

(10) HAIR (on the head; meaning 'head' is not original): PSa **q'*^{W}*um* as in: IS:MC *q'*^{W}*um-qən* 'head' (='hair + head'; cf.Th lex. suff. *=qin* 'head') : MS *q'*^{W}*om-qən* / CS:Sq *s-q'*^{W}*um-ay* 'hair', etc. // NC **g̱'(w)ămʔə̄* 'plait, mane; hair'. [NB sound correspondences: Sa **q'*^{W} : NC **g̱'w*, umlaut: Sa **u* : NC **a*]

(11) DEER, etc.: MC *x̣ʷəl(ʕ)* 'buck' : Sp *x̣ʷiʔ* id. / CS:Ls *x̣ʷel* 'deer': Sq *x̣ʷiʔ* // NEC ***Gwăl*ā̄** 'doe, hornless goat'. (On spirantization Sa *x̣ʷ* < **Gw* see below). - Possibly related to Sa **x̣ʷiƛ'* 'mountain goat' (Kuipers), as in Se *s-x̣ʷiƛ'-ay*.

(12) COLD: PSa **cụ/ạl* // NC **č'wErHV*. [Pre-Sa **rH* is indicated by vowel retraction].

(13) WASH: PSa **c'aʕʷ* // NC lex. suff. **=Häç̌wĂ* [Metath. in pre-Sa?].

(14) GROW(TH): PSa **ƛ'ax̣ʷ*, as both in IS:Th and CS:Ls, not just **ƛ'ax̣* (Kuipers); this latter is a root variant with a lost [w]) 'grow(th), old'. - PSa root var. **ƛ'ax̣* appears in IS:CO *ƛ'x̣*, MC *ƛ'əx̣* / CS:Se *ƛ'ax̣-ax̣* *'grown up' > 'old person' // NEC **ƛ-'ōrχwV* 'sprout'. [Sa **a* (<**a*, **o*) in both main branches: IS:CO, CS:Ls (*o* is secondary in Sa languages); note *x̣ʷ* : **rχw*].

(15) HARD: IS:Th *ƛ'əʕ̣ʷ* vs CS:Sq *ƛ'əx̣ʷ* / Ts:UC *ƛ'əx̣ʷ* / BC *ƛ'ax̣ʷ* // NC **Ƚwĕr V*. Note typical transfer of **w* from *L*-type sound to *ʕ/x̣* in pre-Sa: [w] shifts to the right, to a more "comfortable" position, after **r* turns to [ʕ/x̣].

(16) BOY: PSa **t(a)wiH* [not **taw* '(small and) growing up'; Sa has two similar roots: one for 'boy', another for 'small, little'; see below] // NEC **dwirx̣E* 'child, son'.

Both IS:MC and CO show the root *tw'i-* 'boy, child', cf. also CS:Ls *tawixʷ* 'child'. - A different root (with the meaning 'little') is present in MC *t'aʕʷ*, CO *tʔiw-*. Accordingly, Kalispel shows *ł-tətəw'i-t* 'youth, young boy' vs *tew* 'little').

For the meaning 'small, little', cf. NC **t'iHV* (which may be **t'iHU*). Contamination of both roots ('boy' and 'little') seems possible in some Sa languages.

There is a tendency to lump together two PSa roots: **k'Vm(ʔ)-* '(take a) handful', as in Sh *k'm-* (:NEC **k'ēm V* 'armful, handful') and a phonetically similar root **k'əmʔ* which means 'bite'.

There is a tendency to lump together two unrelated PSa roots: **ław* 'leave (behind)' and **łup'* (as in Se; cf. N parallels) 'peel off'.

A reconstruction **p'alan(ʔ)* 'tree bark' (Kuipers) does not reflect a rather archaic structure *CVCCV(n)* of this root as represented by CS:Se *p'el'an*, IS:Th *p'aʔyan* (*y* < *l*, etc; we may reconstruct PSa **p'aHlan* / **p'alHan* 'bark', or the like).

As mentioned above, Sa - NC comparisons seem valuable, even if we deal not with PSa reconstructions but with certain forms that appear in individual Sa languages. When comparing pronouns of the 1st and 2nd p. (most stable elements in any language) we may cite either Sa proto-forms or existing Sa forms: there is practically no difference:

(17) PERSONAL PRONOUNS: 1st sg. *-n* (:NEC **nɨ̆*); *-ca/-s* (:NC **zō*);

2nd sg. *-xʷ* (:NEC **ʁwV̄*); *-w* (:NC **u̯ō);*

1st p. pl. *-ł* 'we' (:NEC **Lā̆*).

[Note that NC **ź̌wĕ* 'you (pl.)' has an exact parallel in Wk **-zu* (related to 2nd sg.). - Note also Ath 2nd pl. (subj.) *-*x̣ʷ-*, which is comparable with Sa *-xʷ* : NEC **ʁwV̄*, 2nd sg.].

The above-mentioned Sa root with the meaning 'two' exists only in one language (MC); still, its comparison with NC reveals some archaic relationship between Sa and NC:

> (18) TWO: IS:MC *tq'aw'-s* 'two' (cf. *t'q'ʷmaw'-s-ən* 'together') // NWC **tq/ ':ʷA* (> Ubykh *ṭq̇ʷa*) (:Kartv[elian] **t'q'ub* 'twins', a borrowing?) vs NEC **q'Hwā̆* 'two'. This latter also appears in **q'HămVȽwV* 'one of several wives' (Starostin). We may compare Sa:MC root *t'q'ʷmaw'-* (in 'together', above) with NC **tq'Hwă-mVȽwV*, on which the above NEC **q'HămVȽwV* is based. - Cf. Tsez *q̇/uƛu* 'pitchfork' with the uvular (not velar) initial, possibly influenced by *q̇/a-no* 'two' (Starostin) (:Sa in Ts:UC *q'ʷəx̣* 'fork; split, divide'; *q'ʷəyəƛ'* 'cut in two' ?); see NWk *q'- ƛh-* in ex. 19.

We may consider Sa numerals 'two' (above) and 'three' (next) as genuine proto-language inheritance since both these numerals have parallels in NC. (Some synonymous numerals in Sa may have been borrowings from Penutian, which belonged, along with Sa, to an old North American *Sprachbund*).

> (19) THREE: CS:Ls: *ɬixʷ* // NC **ƛHĕ* (:NWk: Ha *q'-ƛh-* 'six', *'two triads' (?), see ex. 18). - Sa frequently reshapes underlying roots as *CVC*.

> [For root structure *CVC* in Sa vs a different, older structure in NC, cf. also: a) HEAR: CS:Ls *luh* // NC **=eƛu* > NWC **λʷə-*; b) CS:Ls *ʒixʷ* 'first // NEC **c̱wi* 'in front, before'; c) CS:Ls *x̣ʷu-l* 'near' // NC *x̱wE* 'together, close to'; d) IS:MS *xʷuy* 'go' // NEC lex. suff. **=iχ̱wV* 'go'; e) IS:Sp *k'ʷiʔ* 'burnt' (etc.) // NEC lex. suff. **=īk'wV* 'burn, set on fire', - etc.].

It may be rational to compile lists of comparisons between representatives of various Sa language groups (such as Ts, IS, CS) and NC. At some point, we may add Wk cognates to our sets.

We deal with stable roots, many of which are not subject to borrowing; besides, any plausible Sa-NC or Wk-NC cognate set is of interest to us. This short comparison may confirm our thesis that Sa languages are very stable from the standpoint of historical phonetics and that Sa (and Wk) languages easily reveal deep genetic ties with NC languages.

Note that we deal almost exclusively with words/roots that have the same meaning both in Sa (also in Wk) and in NC. [NC data are from the above-mentioned NCED (with a few corrections from Starostin's materials as presented at the Moscow 2000 Conference on deep reconstruction)].

> ANGER/ANGRY, ANT, ARROW, BEND, BLACK, BLOOD/BLEED, BLUE, BONE, BOY, BRANCH, BREAK, BURN, BUTTOCKS, CHILD, CHIP, CLOSE (adverb), COLD, CRAWL, CROWD, DARK, DEW, DIRT, DOG, DRINK, EAGLE, EAR, EYE, EYEBROW, FAT, FLASH (verb), FLOW, FRESH, FOREHEAD, GREASE, GROW, HAIR (on the head), HAND, HANDFUL, HARD, HEAD, HIDE (noun), HORN, I, JOINT, LEAF, LEG, LITTLE, MAKE, MALE, MILK, MOUNTAIN GOAT, MOUTH, OLD, OPEN (verb), POINT (verb), PULL OUT, QUICK, RETURN, RIPE, ROCK, RUB, SCATTER, SCOOP, SCRATCH, SEARCH, SEW, SHARP, SHARPEN, SHORT, SHOULDER, SKIN, SLIP, SNOW, SPEAK, STICK (noun), STINGING INSECT, SWALLOW (verb), THOU, THREE, THROW, TIE, TREE, TURN AROUND, TWO, WAR, WARM, WASH, WE, WEAVE, WHITE, WOMAN, YOU (pl.).

II. A Short Preliminary List of Sa – NC Cognate Sets

The following comparison covers three groups of languages: 1) Ts:UC; 2) IS:MC / MS / Th; 3) CS:Ls / Se. The order of the first (and the second) consonants in Sa roots (capital cons. = Cons. class, for instance, *Q = q q'* q^w q'^w; *X = x ẋ* x^w $ẋ^w$ *ɣ ʕ* $ʕ^w$; *S = s z*; *L = ƛ' ɬ l l'*, etc.) is as follows: *K Q X ʔ // C S T n // r L y // P m w*.

1. A Few Comparisons Between Sa:Ts (represented by UC) and NC Languages

[Ts:UC vs NC]

(1) BURN: UC *k*$^{(w)}$*əw-* // NEC lexical suffix **=ōgwV*. [Sa k^w < *k^w or *g^w].

(2) BE AFRAID: UC q^w*anu-* // NEC lex. suff. **=Hä-GwVn*. [Sa q^w < *q^w or *G^w].

(3) DRINK: UC q^w*o·ʔ* (from PSa; cf. Th *ʔu-*q^w*eʔ*, Ls q^w*uʔ*, etc.) // NEC **ʔu-qwV*.

(4) BARK: UC q'^w*iɬ-* 'cedar bark' // NEC **q'wăɬV* 'bark'.

(5) (?) ARROW: UC *ẋəlẋaʔs* (< **ẋəlʔa-* ?) // NEC **ɦwăɦlV* (also *-ɬ-*) (> Tsez. **heɬ*).

(6) SPEAK: UC *ẋəw-əq'-* // NEC **=ĭχwA* (lex. sf.). [Note Sa *ẋew-* : NC **-χw-*; cf. ex.1 above].

(7) DIRT: UC *ciq*^{w}*-i-* // NC **č̣'Hɨq̄'wĂ*.

(8) SUCKLE (etc.): UC *c'am-i-* 's.' (<PSa **c'ạm'*) // NEC **=ĕč'Vm-* 'gnaw, chew' (AvA **č̣am-*)

(9) DEW: UC *səx*^{w} 'wet, dew' (from PSa **sVx*^{w}) // NC **šaχwV*.

(10) HORSE, DONKEY: UC *tiqiw'-* 'h.' (from PSa) // NEC (Lak.-Darg.-Lezg.) **dHogwā* 'd.'. [This word seems to originate from NC **t'HōgwV* 'hoofed animal'; cf. related NC **t'ūgV* '(young) male goat' (also 'young animal' in NWC].

(11) LEAF, PLANT: UC *ƛ'əc'-* 'grow' (of plants) [*ə* <**a* ?] // NEC **ƛ'ać/ča* (*/-ĕ*) 'leaf' ('plant' in some lang.) (Cf. NEC: Lezg. **ƛ̣ač̣a* 'stem, stalk, leaf, grain').

(12) LOOK (FOR): UC *ƛ'i* 'evidently' (:MC *ƛ'aʔ* and Sp *ƛ'eʔ* 'look for') // NEC *ʔiƚ̱V* 'look'.

(13) (?) DEER (etc.): UC *ƛ'alaš* 'deer' // NEC **ƛ̱'äɦɬə̄* 'lamb'. [This latter doesn't match N **t'älV* 'young (of animals)' (St. '89, no. 197); for N, cf. Sa:CS:Ls *(s-)t'i-t'ələʔ* 'young (fawn, calf, colt)'].

(14) WOMAN: UC *ɬanay'*, lex. suff. *=ɬn(?)* [:CS:Ls *ɬadeyʔ* (*d<n*)] // NEC (Darg.-Lezg.) **λɨnɦV*.

(15) STEAM: UC *pəx*^{w} (:Th *pəẋ*^{w} 'spray with mouth') // NC **pHǟχV*. [NC *-V = -U* ?].

(16) HANDFUL: UC *mo·ʔi-* '(take a) handful' (cf. *mo·-mɬ* 'take a handful'; *-mɬ* to NEC **mă̆r[λ]o* 'handful') // NEC (Tsez.-Lezg.) **mHŏχɨ* > Tsez. **mɔχV*

(17) PAY: UC *muẋ*^{w}*i* // NC **mVx̱wV̆* 'price, pay' (> AvA **mix*^{w}*V* 'pay').

2. A Few Comparisons Between Sa:IS and NC Languages

[IS:MC vs NC]

(1) HAND: (?) MC _kalx_ // NEC _*kwiʡi._

(2) BONE(S): MC _k'_^{W}_ən'_ 'bones for stick game' (root) // NC _*k'(w)inV_ 'small bone'.

(3) HORN: MC _qəx̣-min_ (root + instrum. suff.) // NC _*qwĩrhV._ [Late delab. in Sa; cf. NEC:Lak _qi_ 'horn'].

(4) CRAWL: MC _q'əw'-t_ // NC _*HV-q̱(w)V._ [Note Sa _CVw-_ vs NC _-Cw-_].

(5) MALE: MC _s-x̣al-wiʔ_ 'husband' // NC _*χōlʔi_ 'male' (human/animal).

(6) BONE: MC _c'ạm'_ (root; from PSa) // NEC _*Hc̱'wējmə̆_ 'leg bone'. [:Eyak-Ath _*c'émʔ_]

(7) TREE: MC _c'əl_ (root) // NWC _*c̣əla_ 'tree' < NC _*c̱'ə̆/ŏɬV_ 'branch, tree'.

(8) STAND UP: MC _c'əlix_ (not related to ex. 7) // NEC _*=Vm-ć̱'Vr._

(9) DARK: MC _c'el'_ 'shadow, dark', _c'al'_ 'shady' // NEC _*Hč'ōlV̄_ 'black' (Lezg. _*c'olV_ 'black; dark berry' = raspberry, etc.).

(10) TWO: IS:MC _tq'aw'-s_ (cf. _t'q'_^{W}_maw'-s-ən_ 'together') // NC _*tq'Hwā̆_ > NWC _*tq/':_^{W}_A_ (> Ubykh _ṭq̇_^{W}_a_).

(11) (?) ROT, PUS: MC _naʔq'_ 'rotten meat' // NEC _*nĕwq'ŭ_ 'pus' (> Lezg. _*näwq:_) [:ST _*n0k/ŋ_].

(12) BREAK (etc.): MC _ləq'_^{W}_-_ 'break, smash' // NEC _*Hlŏq̱V̄n-_ 'destroy, break, scatter'. [Note Sa _-q'_^{W} vs NC _*-q̱V_, possibly _*-q̱U_ ?].

(13) (?) (TELL) A STORY: MC _may'x_ (root) 'tell a story' // NEC _*mħ/ʕārχwā_ 'tale'.

[IS:MS vs NC]

(14) PLENTY: MS _q_^{W}_o/ey_ 'rich, plenty' // NC _*q̱'(w) ă/ĕjē_ 'things, possessions'.

(15) GO: MS _x_^{W}_uy_ 'go' // NEC lex. suff. _*=iχ̱wV._

(16) SPARK: MS _cʔik_^{W} (:_cik_^{W} 'shine') // NWC _*c̣V(jə)ḳ_^{W}_a_ < NC _*c̱'wɨ̆jVkwŏ_ 'brand, spark'.

(17) SHEEP, LAMB: MS _ɬox̣_^{W} 'sh.' (:Sp _ɬuʔ_) NC _*ʕĩlχU._

[IS:Th vs NC]

(18) DRINK: IS:Th _ʔu-q_^{W}_eʔ_ (:Ts:UC _q_^{W}_oʔ_) // NEC _*ʔuq̱wV_ or _*ʔōq̱wV._

(19) JOINT: Th _q'_^{W}_ət-x̣_^{W}_əm'_ (2nd stem: 'lump') // NC _*q'HwəntV_ 'knee, elbow'.

(20) (?) ELK, GOAT (etc.): Th _təx̣_^{W}_aq'_^{W}_-_ (root) in _təx̣_^{W}_aq'_^{W}_-iʔpe_ 'doe' // NC _*dVrq'wV_ 'he-goat'.

(21) BEND: Th _ɬaq'_^{W}_-əw-t_ 'bend over' // NC _*=ilq̱'wV(ɬ)._

(22) ROCK: Th *ƛ'ix̣*^{w}- // NEC **ƛ̱'wehrŭ* (or *-lŭ*). [*w*-transfer in pre-Sa?]

(23) GATHER; HANDFUL: Th *muq*, *moq*^{w}- 'gather' // NEC **mŏg̣'V̌* 'handful' (also 'handle, hilt') (> Tsez. **moq*).

(24) SWALLOW, THROAT (etc.): Th *məq'* 'satiate', *məq*^{w} 'hold in mouth' (:Ls *bəq'* 'put/hold in mouth, swallow'; *b* < *m*) // NEC **mVq̱'VɬV* 'throat, larynx' (> Tsez. **muq'* 'throat').

(25) WHITTLE, etc.: Th *məx* 'wh.; sliver' // NEC **m[ā̆]nx̱wV* 'sickle' > Lak *mirx* (etc.). [NC also shows meanings 'plough', 'reap' in different languages].

(26) SNOW: Th *məx̣*^{w} (:Sp *mx̣*^{w}- 'to snow') // NEC **marχalV* > Lezg.:Tab. *maʁ*^{w}*al.*

(27) GREASE: Th *mi*$^{ʕ'w}$- (root) // NEC **măɦwV.* [Voiced cons. both in Sa and NC].

(28) LYNX, CAT: Th *məɣew'* 'l.' // NEC **mHarGVwV* 'tom-cat'. [As above].

(29) STINGING INSECT(S): Th *məz·/mə̣ze* 'flies'; *məc'·/mə̣c'e* 'bees, hornets, wasps' // NC **miʒȀ* stinging insect.

(30) MOUNTAIN, HILL, PILE: Th *mol* 'pile up (dirt or snow)' // NEC **muHalV* 'mountain' > Lezg.: Arch. *mul*, etc. (In Tsez. also 'hillock, knoll').

3. A Few Comparisons Between Sa:CS and NC Languages

[CS:Ls vs NC]

(1) BURN: Ls *k'*^{w}*a-s* 'burned' (cf. IS:Sp. *k'*^{w}*iʔ* 'burnt') // NEC **=ɪ̆k'wV.*

(2) (?) GREEN (etc.): Ls *q*^{w}*ac* 'yellow, (light) green, pale' // NEC **GoʒV* 'green color, dirt'.

(3) THOU: Ls *-ax*^{w} // NEC **ʁwV̄.*

(4) SCATTER: Ls *x̣*^{w}*əš* (*š* from *x*) // NEC **=Hĕχ wV.*

(5) SHARP: Ls *x̣*^{w}*əc* // NC **ɦüǯwĀ* > Tsez. **ʔãč-*. [*w*-transfer (from *C* to *X*) in Sa?].

(6) (?) EAT: Ls *x̣*^{w}*uƛ'* [:NWk *ƛ'x*^{w}-, metath.?] // NEC **=ɪʔwVl* (metath. in Lezg.:Archi *lah-* < **liʔ*^{w}*ä-*).

(7) BLOOD: Ls *cəɬ* // NEC **ć'āŁwV.* [There is no PSa **cay* 'blood' (*y* < **ɬ*); for Sh *ci*^{ʕw}, Th *cʔi*^{ʕw}, 'bleed' (Sq root *ciʔ/x̣*^{w}), cf. Nakh. *c̣ēgi*, Lezg.Tab. *čiwi* (same NEC root)]

(8) CHILD: Ls *tawix*^{w}- (root) // NEC **dwirχE.* [Note Sa *CVC-* vs NC *CC-*].

(9) ACROSS: Ls *t'aɬ* '(put) crossways, across', *t'aɬ=us-ən* 'beam' // NEC **t'wēlʔe* '(cross)beam'.

(10) WE: Ls *-aɬi* // NEC **Lä̃.*

(11) MUCUS, SALIVA: Ls *ɬəbc'* 'm.' (*b* < *m*) // NEC (AvA-Tsez.) **ƛămVč'V* (/*x*-) 'saliva, pus'.

(12) (YOUNG) HORNED ANIMAL: Ls _s-wiʔ-qaʔ_ 'buck deer' (root _*wiʔ_) // NEC _*wɨ̆ɦwV_ 'sheep, lamb, young (horned) animal'.

[CS:Se vs NC]

(13) HIDE, STEAL: Se _kʷal-_ 'h.' // NEC _*=igwVɬ_ 's.'. [Sa _kʷ_ < _*kʷ_ or _*gʷ_].

(14) BEND, CURVE: Se _k'ʷuc'-_ 'bend' (v.) // NEC _*kɨ̆ɛ̆'wɨ̌_ (also _*č'ŭ̄kwV_) 'curved'.

(15) (?) LUMP (body): Se _qʷemx̣ʷ_ 'lump of ankle' // NC _*q'ămq'(w)ă_ 'knee, leg-bone'.

(16) GOAT: Se _x̣ʷiƛ'ay_ 'mountain goat' // _*Gwăɬā_ 'doe, hornless goat' (> AvA _ʁʷalV_ > Tlan. _ʁʷeli_ / Tsez. _*ʁ:ʷel_ > Gin. _ʁʷil_, etc.). [NB: Altern.: NC _*ʔĭlχU._ (Metath. _*x̣ʷilʔ-_ in pre-Sa?)].

(17) (?) DRIP, DRIZZLE: Se _c'iq'ʷ-_ 'drip' // NEC _*čŏwqɨ̄_ 'drizzle'.

(18) WASH: Se _c'əx̣ʷ_ // NC _*=Häć̱wĂ._ [Metath. and _*w_-transfer in pre-Sa?].

(19) LOOK, EYE: Se _c'il-_ 'look for' // NEC _*c'ilV_ 'eye' [sic! Not 'eyelash, eyebrow'].

(20) GIVE: Se _yat-_ (from _*=it-_ ?) // NC _*=VtV̌._

III. Selected Wk - NC Cognate Sets

Phonetically, Wk languages are more archaic than Sa [a higher frequency of laterals (close to that of NC); a higher frequency of inherited voiced consonants, etc.). We may use the following data.

(1) WIDE: Wk:No, Ni _ʔaq-_, NWk:He _[ʔʌqá]_ // NC _*ɦărq'Ḗ._ [_-q-_ < _RC_, as in Sa].

(2) HOLE: Wk:Ni _kux̣ʷ-ak_ // NEC _*kHwə̄rV._ [Wk _x̣_ < _*r_, as in Sa?].

(3) KNIFE, SCRAPE: NWk _kus-_ 'scrape off with a knife, shave' // NEC _*ḵ[ĭ]śwV_ (cutting tool). [Transfer of _*w_ in pre-Wk ?].

(4) SEE; VISIBLE: Wk:No _kʷa·hi_, M. (=Ma?) _kʷa·x̣i_ 'v.' // NEC lex. suff. _*=agwV_ 'see'.

(5) DUST, DIRT: NWk: _q'x̣ʷ_ 'dust, powder' // NEC _*q'ħ/ʕwŏrV_ 'dirt, turf'.

(6) NECK, THROAT, GOITRE: NWk:He _q'ʷú-qʷùni_ 'neck' // NEC _*q'winV_ 'goitre, Adam's apple'.

(7) BUTTOCK(S): NWk:Kw _x̣ím'a_ // NEC _*χɦwɨ̆mχV_ (also 'cheek') > Lezg. _*χ/ɨ(m)χ/-_ 'buttock'.

(8) DIRT: NWk _cqʷ-_, _zqʷ-_; _c'qʷ-_ 'dirty'; cf. Kw _ʒeqʷá_ 'dirt' [:Sa:UC _ciqʷ-i-_ 'dirt'?] // NC _*č̱'Hɨ̆q'wĂ_ (_/-G-_).

(9) SLIP: NWk _caʔx-_ (as in Kw _ceʔxá_ 'slippery') // NEC _*čirχwV._ [Delab. in pre-Wk?].

(10) (?) FLOW: NWk:Kw _cax̣is_ 'flowing down' // NEC _*čHaχV._

(11) FLOW, POUR; WASH: NWk _cxʷ-_ (as in Kw _cxʷla_ 'overflowing') may, or may not, relate to NWk _c'ux̣ʷ-_ > Kw _c'ux̣ʷa_ 'wash' [cf. in Sa: UC _c'əx̣ʷ-_ and MC _c'aw'_ 'wash' vs MC _c'əx̣ʷ_ 'pour out'] // NC lex. suff. _*=Hä-ć̱wV_ 'pour, wash'.

(12) FAT: NWK *cnx̣^{w}-* // NEC **c'ēnx̱wV*

(13) SHAKE: NWk:Kw *culix̣a* 'shatter' // NC **=ēʒwEl.*

(14) SHORT: NWk *c'k^{w}-*, Kw [*c'ek'^{w}ā*] // NC *ǯĭk'wV̆.*

(15) SCOOP: NWk *c'iq-* // NEC **c̱ǎ̆q'wă.* [Cf. next, for Wk *q* (etc.) vs NC *q'w*].

(16) NARROW, TIGHT: NWk *c'iq'-*, *c'iq'^{w}-* (also with *q*, *q^{w}*) // NEC **č'iq'wV* 'tight plait'.

(17) (?) POLE, STICK: NWk *c'wax̣-*, Kw *c'ux^{w}-* 'insert (pole)', *c'x^{w}-* 'stab'; *zux^{w}-* 'log, pole' // NEC **c̱'wēx̣V* 'stick'.

(18) ARROW: No *c'iḥat*, Ma *c'ix̣at* (Wk root **c'ix̣-* ?) // NC **c'änHV*

(19) FRESH: NWk:Oo *c'uta* (<**c'w-* ?) // NC **=Vč'wV* 'good, fresh, new' > NWC **č̣^{w}a.*

(20) BLACK: NWk *c'uɫ-* [:Sa:BC *c'u-* 'grey' ?] // NEC **Hč'ōlV.*

(21) (TO) POINT: NWk *c'm-* 'idex finger; to point' [:*c'm-t* 'stand on tiptoes'] // NEC **c̱'ūmV* 'tip, point'.

(22) PLANT: NWk *zm-* (Kw *zmiʔ*, etc.) // NEC **č'ā̆mħV.*

(23) SHARPEN: NWk **zux̣^{w}-* // NC **ɦüǯwĀ.* [Metath. and *w*-shift in pre-Wk?].

(24) MILK: NWk: Kw, Oo *zam'-a* '(suck at the) breast'; **zm-x^{w}-* 'milky' (in Oo) // NC **=āmʒŬ*

'to milk, to drink'. [Wk underlying root **zm-* from metathesized **zVmV* ?].

(25) CROOKED, CURVED: NWk:Oo *siʔk-la* 'crook-ed' // NEC **č'ü̆kwV* 'be hooked, curved; cook'

(26) (?) MOUTH: NWk *sm-s-*; Oo *sm-yat* 'have mouth' // NEC **ʒ́wĕmV.* [For phonetics, cf. Wk *-s/zu* 'you' (pl.) : NC *žwĕ* id.; Sa *-s/ca* 'I' : NC *zō* id. ?].

(27) CUT: NWk *t'ew-* (in Kw verb *t'ewik^{w}*) // NC **=āt'wV.* [Wk *CVC* vs NC *CC*].

(28) I: Kw *-en*; cf. *nus*, He *nís* 'be mine' [cf. Sa] // NC **nɨ̌*

(29) DIRT: NWk *niq^{w}-* (Kw, Oo *niq^{w}a* 'dirty') // NEC **nĕq'wɨ*

(30) THROW: NWk:Kw *nep'-íd* // NEC **l(H)ap'V* / Urartu *nāp'-* (as in *nāp'-aχ-* 'overthrow'). [Wk *n* < **l(H)* ?].

(31) SHOULDER: NWk *n'ik^{w}-* 'carry on the sh.' [:NWk *nix̣-* 'act with hand'?] // NC **nHɨ̆wG̱Ā* 'arm, shoulder'.

(32) (?) BERRY: NWk *n'ux^{w}-* (Kw *n'ux^{w}a* 'small blueberry') // NEC **niwG̱V.*

(33) SWALLOW (verb): NWk *n'q^{w}-* (Kw *n'q^{w}a*) // NEC **=HV-q̱'wVn.* [Metath. in Wk?].

(34) (?)MOUNTAIN GOAT: NWk *n'ax̣-* (He *n'ax̣a*) // NEC **ɦ/ʕānx̱V.* [NB Wk *CVC* : NC *-CC-*].

(35) FLASH: NWk:Ha *l'pa* 'flash a light' // NEC **l/ɬapV* 'glitter, flash'.

(36) BRIGHT: NWk:Kw *ɬis-a* 'fair (complexion)' // NC **lōʒV̄* 'bright metal > Lezg. **lac:V-* 'white', etc. (Lezg. languages shift the original meaning 'bright metal' to 'bright, white').

(37) (?) KIN: NWk:Kw *λul'iʔ* 'nephew, cousin' (etc.) // NC *Hĭ-ƛ̱'īwV-ɬV* (> NWC **ƛ':*^{W}*əlA* 'seed, kin, clan, people').

(38) HARD: NWk: He *ƛ'áx̣-* and Kw (etc.) *ƛ'ax̣-* 'stiff, regid' [:Sa:UC *ƛ'əʕ*^{W} / MC *yəʕ'*^{W} hard'] // NEC **ƛ̱wĕrV* 'hard'. [**w*-loss in Wk; **w*-transfer in pre-Sa?].

(39) (?) SKIN: NWk *ƛ'is-* // NEC **ƛ̱wājc'ā* 'skin, bark'.

(40) (ACT WITH) HAND: NWk:Kw *ƛ'ol-* 'feel, grope with hand': NEC **ƛ'ōrə* 'hand'.

(41) DOG: NWk:Kw *w'ac'* // NC **gwăǯē* > Darg. **k:*^{W}*ač:a* (etc.).

IV. Direct Comparison Between Sa and NC Daughter Languages

Since the Sa proto-language is not yet sufficiently reconstructed, and since the existing reconstructions are fully, or almost fully, identical to roots in individual Sa languages (see above), we may *compare roots of individual Sa languages* *directly* *to NC daughter languages* (first of all, to reconstructed languages - ancestors of NC language groups). Predictably, roots/words of individual Sa languages are, on many occasions, identical or almost identical to roots/words in NC daughter languages.

If we consider the following three cognate sets we may see that the closest link between Sa and NC languages is neither Sa-to-NC nor Sa-to-NEC but that between Sa and a certain NC daughter language, such as Nakh:

> (1) EYE: Ts:UC *co·-* (as in *co·q*^{W}*a* 'tear' = *'eye+water') / CS:Se *ca-* // Nakh. **c̣a-* (< NEC **c'alV* > Darg. **-č̣ali*, with preservation of the NEC **-l-*).
>
> (2) DOWN: Ts:UC *lix*^{W}*i-* 'walk downward' // Nakh. **laχu-n* (cf. Chechen *loχa* 'low', *laχa* 'down, below'; *laχ-alar* 'descend, come down', etc.) < NEC (Nakh.-Lak.) **l/ɬV̆χV* 'low'. [Note Sa *-x*^{W}*-* : Nakh. **-χu-*].
>
> (3) GATHER, HANDFUL: IS:Th *muq*, *moq*^{W} 'gather' // Nakh. **muq̇* (< NC **mŏg'V̆* 'handful, handle') / Tsez. **moqV.*

Now we may look through four short lists of cognate sets, namely, Sa vs AvA; Sa vs Tsez.; Sa vs Lezg.; Sa vs NWC. Occasionally, Wk - NC cognate sets are listed as well.

1. A Few Comparisons Between Sa and Avar-Andian Languages

The following list includes some comparisons between Sa languages and NC daughter languages. On occasion, Wk-to-NC cognate sets also appear in the list.

> (1) EYES, LOOK: / Ts:UC lex. suffixes *=us-i-* 'face, eye', *-al-us* 'eye' (:CS:Sq *=us* 'eyes' in *c'ip'=us* 'shut the eyes') // AvA **-us-* 'to search' (> *-us-* in modern AvA l-ges) < NEC **=HimsV* 'to look'. - The above comparison may be summarized as follows:

[=1] Sq *=us* 'eyes', UC *=us(-i)-*	And. *-us-* < AvA **-us-* 'search'	< NEC **=HimsV* 'to look'

(2) TO SCRATCH: IS:Th and MC *qəs* // AvA **qas:-* < NEC **qāl̠s̱V.*

(3) WHITE, BLUE: Ts:UC *q*^{W}*a/uẋ*^{W}- 'wh.', *q'*^{W}*iẋ* 'b.' // And. *q:*^{W}*oj* 'b.' (<PAnd. **ʔV-q:*^{W}*o-ji-* < AvA **q:oʔi-* 'wh., b.' < NEC **=ĕq̱wĂ* 'yellow').

(4) COLLAR-BONE: IS:Sp *-q'wl-* (root) // AvA **q*^{W}*ilu* (< NEC **qHwŏƚwV* 'neck, collar'). [Note Sa *q'(*^{W}*)* vs NC **qH(w)*].

(5) LUMP, BUTTOCK(S): Ts:UC *ẋ*^{W}*um-* 'lump' (in compounds, meaning 'elbow', buttocks'; cf. also IS:Th *q'*^{W}*ət-ẋ*^{W}*əm'* 'joint' vs NEC **q'HwəntV* 'knee/elbow') // Wk:Kw *ẋím'a* [note delab. *ẋ*] // AvA **χ:*^{W}*[i]m(V)χV* (:Lezg. **χ/ɨ(m)χ/-*) < NEC **χɦwɨ̆nχV* 'buttock, cheek'. [NB **-mχ-* > *-*-mVχ-*].

(6) FAT: NWk *cnẋ*^{W}- (in Kw *cnẋ*^{W}*iʔ*, etc.) // NEC (AvA-Darg.) **c'ēnx̱wV.*

(7) SUCKLE, CHEW: (?) Ts:UC *c'am-i-* 's.' (<PSa **c'ạm'*) // AvA **č̣am-* < NEC **=ĕč'Vm-* 'gnaw, chew'.

(8) POUR, WASH: IS:MC *c'əx*^{W} 'pour', *c'aw'* 'wash' : Sp *caw* (*:caʕ*^{W}), *c'ew'* 'wash' // AvA **č:Vb-* (*b* < **w*) < NC **=Hä-ç̱wĂ.*

(9) GROW(TH): IS:Th, CS:Ls *ƛ'aẋ*^{W} // AvA **ƛ̣iχ:*^{W}*i* < NEC **ƛ'ōrχwV.* [NB absence of pre-consonantal **r* both in Th, Ls, - and in AvA; note also preservation of labial elements in these languages].

(10) LOOK FOR: IS:Sp *ƛ'eʔ* / Ts:UC *ƛ'aʔ* (etc.) (from PSa) // AvA *ƛ̣:V-* < NEC **ʔiƚ̱V* 'to look'. [Metathesis in pre-Sa?].

(11) MUCUS, SALIVA: CS:Ls *ƚəbc'* 'm.' (*b* < *m*) // AvA **ƛ/λaç̣:V* and **ƛ/λamç̣V* < NEC (AvA-Tsez.) **ƛămVč'V* (*/x-*) 'saliva, pus'.

(12) PAY: Ts:UC *mux*^{W}*i-* // AvA **mix*^{W}*V* 'pay' (> Av. *mux*) < NC **mV̄xwV̆* 'price, pay'.

2. A Few Comparisons Between Sa and Tsezian Languages

(1) SMOKE: Ts:UC *q*^{W}*o·ʔ* and *q'*^{W}*uẋ*^{W}-; CS:Ls *q*^{W}*əš* (*š* < *x*) // Tsez. **q*^{W}*ɨ̆ *qo* < NC **k̠wɨ̆nħV.* [Sa *q*^{W} matches Tsez. **q*^{W}].

(2) JOINT: Ts:UC *ẋ*^{W}*ut'* 'bent up' / IS:Th *q'*^{W}*ət-ẋ*^{W}*əm'* // Tsez. **q̇ɔtV* and **q̇ɔntV* < NEC **q'HwəntV* 'knee/elbow'. [Note Sa *-t'* : Tsez. *-t-* and *-nt-* < NC **-nt-*].

(3) STICK (noun): Ts:UC *ẋəc'* 'stick' / IS:Th *ẋəc* and *ẋic* 'wooden' // Tsez.:Gin. *χišu* < NEC **GHwälćV* 'stick, board'.

(4) ARROW: Ts:UC *ẋəlẋaʔ-s* (< **ẋəlʔa-* ?) // NEC **ɦwăɦlV* (also *-ƚ-*) > Tsez. **hel* / Lezg. **ħäl(:)* (etc.). [Sa shows *ẋ* vs **ɦw* (similar: NEC languages Tsez. and Lezg.). - Sa also seems to show *l ... ʔ* (from **ʔ ... l* ?) vs cluster **ɦl* in NEC].

(5) SHARP: CS:Ls *ẋ*^{W}*əc* // Tsez. **ʔãč-* < NC **ɦüǯwĀ̄.* [*w*-transfer (from *C* to *X*) in pre-Sa?].

(6) GROW(TH): IS:MC *ƛ'əx̣-*, CO *ƛ'x̣* / CS:Se *ƛ'ax̣-* (etc.) / Tsez. **ƛ̣eχ(:)* < NEC **ƛ'ōrχwV.* [NB preservation of [x̣] but loss of [w] both in CO, Se, - and in Tsez.].

(7) GATHER, HANDFUL: IS:Th *muq*, *moq*^{W}*w-* 'g.' // Tsez. **moqV* (:Nakh. **muq̇*) < NC **mŏg'V̆* 'h.' (also 'handle').

(8) SWALLOW, THROAT (etc.): IS:Th *məq'* 'satiate', *məq*^{W} 'hold in mouth' (:Ls *bəq'* 'put/hold in mouth, swallow'; *b* < *m*); from PSa // Tsez. **muq̇* 'throat' < NEC **mVg'VɬV* 'throat, larynx'.

(9) HANDFUL: UC *mo·ʔi-* // Tsez. **mɔχV* < NEC (Tsez.-Lezg.) **mHŏχɨ-*

(10) HANDFUL: UC *mo·ɬ* //. NEC **mă̌r[ƛ]ə/o* 'handful, armful' (> Tsez.:Gin. *meƛu* 'handful', etc.).

3. A Few Comparisons Between Sa and Lezgian Languages

(1) ANGER: Ts:UC *qəɬəx̣-* 'angry' // Lezg. **q/al(:)* > Lezgi *qel* (:Nakh. **qēl*) < NEC **Gwāłħo* 'gossip, offence, anger' (etc.). [Sa seems to show simplified *q* vs NEC **Gw* (similar: Nakh.; Lezgi). - Besides, Sa shows a typical restructuring of the *CVCCV*-type root into *CVCVC*].

(2) DAY: Ts:UC *q'ix̣-* // NEC **Hwīq̱'V* > Lezg. **q:i* 'today' < **Hwq̱'ī* . [Metath. in pre-Sa?]:

[=2] UC *q'ix̣-* 'day'	Lezg. **q:i* 'today' < **Hwq̱'ī*	< NEC **Hwīq̱'V* 'day'

(3) DRINK: Ts:UC *q*^{W}*oʔ* / IS:Th *ʔu-q*^{W}*eʔ* / CS:Se *q*^{W}*u-* // Lezgi *q*^{W}*a-* < Lezg. **ʔoq*^{W}*a* 'suck, drink' < NEC **ʔuq̱wV* or **ʔōq̱wV* 'drink'; actually a Lezg.-Tsez. isogloss (cf. also NEC lex. suff. **=VqV* 'suck').

(4) (?) LOOK: Ts:UC *ʔəx̣-* 'see' // Lezg. **ʔ*^{W}*Vr* (> Arch. *hara-*, in compounds; cf AvA **ħa/orV*) / Darg. **ħer* < NC **ʔwerV*, noun and verb (an archaic root).

(5) DARK: IS:MC *c'el'* 'shadow, dark' // Lezg. **c'olV* 'black; dark berry' (= raspberry, etc.) < NEC **Hč'ōlV̄* 'black'.

(6) EYE: CS:Se *c'il-* 'look for' // Lezg. **c̣il-* in **c̣il-c̣im* [sic!] 'eyelash' > Tab[asaran] *c̣il-c̣im* / Tsez. **c̣ɨ(l)-c̣(im)-* id. (to NEC **c'ĭlV* 'eye'). - Se (Sa) and Lezg. (NEC) show an archaic feature: preservation of the 2nd root cons. **l.*

(7) BLINK THE EYES / EYELASH: BC *c'im-uɬ* 'blink the eyes' [:UC *c'im-alis-* 'open and shut the eyes; squint'] // Lezg. **c̣il-c̣im* > Tab. *c̣il-c̣im* 'eyelash' (:Drav. **cimV* 'blink, wink; eyelash'). - See above.

(8) (?) ROT(TEN), PUS: IS:MC *naʔq'* 'rotten meat' // Lezg. **näwq* > Lezgi *naʁ*^{W} (:Tsakh. *naʁ*) < NEC **nĕwq'ŭ* 'pus'.

(9) LEAF, PLANT: Ts:UC *ƛ'əc'-* 'grow' (plants) [*ə* <**a* ?] // Lezg. **ƛ̣ač̣a* 'stem, stalk, leaf, grain' < NEC **ƛ'ać/ča* (/-*ĕ*) 'leaf' ('plant' in some lang.) [Lezg. glottalization pattern equals Sa].

(10) WOMAN: Ts:UC *łanay',* lex. suff. *=łn(?)* / Cs:Ls *ładəy?* (*d* < *n*) // Lezg. **λ:ɨn:(ol)* < NEC (Darg.-Lezg.) **λɨnɦV.*

[Sa:UC word seems to match Lezg. root precisely: UC *-ay-* matches Lezg. **-ol,* etc.].

(11) SNOW: IS:Th and MC *məx̣*^{W}*-* : Sp *mx̣(ú)* 'to snow' // Lezg.:Tab. *maʁ*^{W}*al* < NEC **marx̱al/łV.* (12) MOUNTAIN, HILL, PILE: IS:Th *mol* 'pile up (dirt or snow)' // Lezg.:Arch. *mul* < Lezg. **muh/ Vl* or **muʔ/ Vl* < NEC **muHalV* 'mountain', etc. (In Tsez. also 'hillock, knoll').

(14) EYE: CS:Se *wil/'-* 'peep, peer', BC *-uł* // Lezg. **ʔ*^{W}*il* > Lezgi *wil,* Tab. *ul* / Darg. **ħuli.*

4. A Few Comparisons Between Sa and NWC Languages

(1) STICK (etc.): CS:Ls *q*^{W}*(ə)łay?* 'stick, log' // NC **q'wălʔă* 'board' > Lezg. **q̣ula* 'board, shelf, lid, small plank' / NWC **G*^{W}*ə* 'board, post, pole, stake'. [Non-glott. Sa *q*^{W} matches NWC **G*^{W}].

(2) SPARK: IS:MS *cʔik*^{W} (:*cik*^{W} 'shine'; cf. Th,Sp *c'ek'*^{W} 'shine, shiny') // NWC **c̣Vḳ*^{W}*a* / **c̣Vjəḳ*^{W}*a* < NC **c̱'wɨ̆jVkwŏ* 'brand, spark, brilliance':

[=2] Sp *c'ek'*^{W} 'shiny'	Tsez. **c̣əḳə* 'fire-brand'	WC **c̣Vḳ*^{W}*a* 'spark, fb'	< NC **c̱'wɨ̆jVkwŏ*

(3) FRESH: NWk:Oo *c'uta* (<**c'w-* ?) // NWC **č̣*^{W}*a* < NC **=Vč'wV* 'good, fresh, new'.

(4) TWO: IS:MC *tq'aw'-s* (also *t'q'*^{W}*-* in this root) // NWC **tq/':*^{W}*A* > Ubykh *t'q'*^{W}*a.*

(5) DO, MAKE: Ts:UC *wi* // NWC **wə* < NC **=ăhwV(r)* [ablaut **a/i*] (> NEC:Tsez. **=Vw-*).

V. Genetically Related Compounds in Salishan and North Caucasian Languages

We may identify CS:Se *ca-cum(-an)* 'eyebrow' with Nakh. **c̣a-c̣ʔVm* 'eyebrow'; this compound apparently consists of a word for 'eye' + word for 'hair' (:ST **chām*): this latter component appears as **-c'ɦwĕme* in NEC (in 2 compounds, both for 'eyebrow' <'eye+hair'):

(1) EYEBROW: CSSe *ca-cum-an* / Ts:UC *cum-ay'is* (*'hair' + 'eye') // (NEC **c'alV-c'ɦwĕme* [rather than **c'ilV-c'ɦwĕme*] >) Nakh. **c̣a-c̣ʔVm* 'eyebrow'; cf. next ex. where the component 'eye' is different.

(2) EYEBROW: BC *-uł* 'eye' + Se, UC *cum* in 'eyebrow' (ex. 1) // (NEC **ʔwiłʔi-*c'ɦwĕme* >) Lezg. **ʔwil(i)-c̣*^{w}*em* > Tab. *ul-č̣*^{v}*i/am* (eye+*hair). Cf. 'eye': (Lezg.>) Lezgi *wil*; (Darg.>) Ak. *ħuli.*

The NEC component **c'alV* and/or **c'ilV* means only 'eye': gloss "**c'ilV* 'eyelash'" in NCED seems incorrect: when this root appears in compounds with the meaning 'eyelash' *then the 2nd* *component is genetically different from the above* **-c'ɦwĕme* (> Nakh. **-c̣ʔ Vm*, Lezg. **-c̣ʷem* etc.)

Accordingly, *in a synonymous compound* **ʔwiƛ̄ʔi-*c'ɦwĕme* *the 1st component means only 'eye', not 'eyelash'*. The 2nd component (*-*c'ɦwĕme*) *is used only in words for 'eyebrow', not in words for 'eyelash' (or 'eyelid')* [cf. Sa:UC *cum ay'is* 'eyebrow' (*'hair' + 'eyes'): not to *c'im-* 'blink'].

As we have seen, NEC compound **ʔwiƛ̄ʔi-*c'ɦwĕme* 'eyebrow' > Lezg. **ʔʷil(i)-c̣ʷem* 'eyebrow' is the underlying form for Tab. *ul-č̣ʷi/am* id.; labial element [w] in the 2nd component indicates that it originates from NEC **-c'ɦwĕme.* In Sa. *wil-/wi'-* (CS:Se) means 'peep, peer', and a related form *-uł* means 'eye(s)' (acc. obj. in BC *c'im-uł*); Sa *cum* is used only in words for 'eyebrow'.

It is incorrect to reconstruct the NEC word for 'eyelash', using the 2nd component derived from NEC **c'ɦwĕme*: this latter always means 'eyebrow', never 'eyelash': so, there is no "**c̣il-c̣⁽ʷ⁾em* 'eyelash'" in Lezg.:Tab. *c̣il-c̣im* 'eyelash'. This root *-c̣im* (Lezg. **-c̣im* > Tab. *-c̣im*) *is etymologically different* from *-č̣ʷi/am* (< Lezg. **-c̣ʷem*) in the word for 'eyebrow': *ul-č̣ʷi/am* above.

[In Sa, *c'im-* means 'blink; contract; open and shut' (+ obj. 'eyes' in BC (*-uł*) and UC (*-alis-*)). In NC we have Lezg. **-c̣im* in Tab *c̣il-c̣im* 'eye-lash', and Darg. **č̣imi-č̣ali-* 'eye-lash' (in Ak.). - In Drav. **cim-* means 'wink, blink, twinkle; eyelash, eyelid'. - In Kartv. **c'am-* means 'instant' [typologically matching Russ. *mig* vs. *migat'*, a verb]; **c'am-c'am* 'eyelash(es)'. Old ablaut **i/a* in N **c'imV/*c'amV* [Dolg.: **ć'-*] 'blink the eyes; eyelid' reminds us of **i/a* in NEC **c'īlV/*c'alV* 'eye'.

We may add that NC shows only nominal meanings for the above words ('eye', 'eyelash' etc.); Sa shows only nominal meanings for *ca-*, *-uł* 'eye(s)' but only verbal for Se *c'il-* ('look for'), *wil/'-* ('peep, peer'). - Drav. (**cim-*) is used both as verbal and nominal stem. - Kartv. **c'am-* means both 'instant' (verbal origin very likely) and 'eyelash' (**c'am-c'am-*; verbal origin likely)].

It is incorrect to assert that words for 'eyelash' in many Lezg. languages "are completely distorted": they rather belong to different roots: Tab. Khiv. *mic̣-mic̣* may originate from the above **c'im-*; Khl. *č̣eṗ-č̣eṗ-aj* 'eye-lash' (from *'cover-cover for/of eye'?), Kryz *c̣äp* id., Fit. *c̣iṗ-c̣iṗ* id. may match IS:MC *cəp- ł -ay'* 'eyelash(es), eyelid'; Ls *c'ip-l-il* 'shut the eyes', CS:Sq *c'ip'-us* id., etc. (Note NEC:Khl. *-aj* vs IS:MC *-ay'*, probably, *'eye, face').

Note also that words of the type Tsez. 'eyebrow, eyelash' **c̣ic̣* > Tsez. and Gin. *c̣e-c̣* (as stated in NCED) seem to contain (**c̣il-* > **c̣i* - >) *c̣e-* *'eye' + the 1st sound of the 2nd component (apparently, < **-Vm-* judging from the oblique base Cez. *c̣e-c̣mo-*). We may add that the 2nd component may be *'hair' in words for 'eyebrow', but in words for 'eyelash' the root may be different (possibly **c̣im(V)-* used in words for 'eyelash'; a homonym).

We have, both in Sa and in NEC

(a) Ts:UC *cō-* 'eye' / CS:Se *ca-* // NEC **c'alV* > Nakh. **ça-* / Darg. *-*ç̌ali* 'eye'

(b) CS:Se *c'il-* 'look for' // (NEC **c'ilV* 'eye' >) Lezg. **çil-* 'eye' / Tsez. **çɨ-* 'eye'

(c) CS:Se *wil-* 'peer' / BC *-uɬ* 'eye' // (NEC **ʔwiɬʔi-* >) Lezg. **ʔʷil(i)-* 'eye' > Tb *ul-* 'eye', etc.

(d): [*'hair' in 'eyebrow'] Ts:UC *cum-* / CS:Se *ca-cum-*, Ls *cub-əd* // Lezg. **-çʷem* in Tab. *ul-ç̌ʷi/am*

(e) BC *c'im-uɬ* 'blink the eyes' / (Lezg. >) Tab. *çil-çim* 'eyelash' : Darg. **ç̌imi-çali* id.

VI. Some Other Roots and Compounds Designating Body Parts

There are many forms in both in Sa and NC, describing body and its functions. (Sa compounds of the type 'elbow-joint' relate to words, meaning 'twist, bend' (etc.), or 'lump' [note Sa variants with *q'ʷ/x̣ʷ*]):

(1) FINGER: IS:MC *=aks-t* // NC **k'ă̌ṣɨ/a* (very archaic)

(2) SMALL BONE: IS:MC *s-k'ʷən'-k'ʷən'* 'bones (for stick-game)' // NC **k'(w)inV* 'small bone' (also about small bones for playing dice).

(3) ARM, EMBRACE (etc.): CS:Sq *q'ac'* [*c'* < **rC'* ?] // Lak. and Khosr. *qaç̌* 'shoulder' < NEC **qărć'wV* 'shoulder, arm'.

(4) HORN: (I) IS:Th *qʷay'* < **qʷa/ir'* (?) [*y* may originate from **l*, **r*] / (II): IS:MC *qəx̣-* , *qx̣-* (+ stressed "instrum." suff. *-min*), Sp and MS *qx̣-* (also with suff.) // NC **qwĩrhV* [// N **kErV* (NB Sa stem II); St. '89 # 86].

(5) BREAST: Wk **x̣u:ɬ* (or sim.) : No *ḥu·(ƛ)* / NC **Gwălħē* 'udder, breast' (> AvA *ʁʷarHV* > Kar. *ʁori*, etc.).

(6) HAIR (on the head): IS:Th *q'ʷum* 'head', MC *q'ʷum-qən* 'head' (*'hair' + 'head'; cf. Samish *qən* 'head', loc.) : MS *q'ʷom-qən* / CS:Sq *s-q'ʷum-ay* 'head hair' // NC **q̱'(w)ămʔə̄* 'plait, mane; hair'. Cf. compounds: Sa:MC *q'ʷum-qən* 'head' < *'hair on the head' vs NEC **q̱'(w)ăm-čV* 'hair on the temples'.

(7) JOINT: IS:Th *q'ʷət-x̣ʷəm'* (*'elbow/knee [=joint]' + *'lump'?) / Ts:UC *x̣ʷut'* 'bent up' // Tsez. **q̇ɔ(n)tV* < NEC **q'HwəntV* 'knee/elbow'. [Note Sa *-t'* : Tsez. *-t-* / *-nt-* : NC **-nt-*]. - See next.

(8) (?) *LUMP: Ts:UC *s-x̣ʷum=ač'a* 'elbow' (*'lump + arm'?), *s-x̣ʷum=nač* 'buttocks, hips' ("*lump + leg'?) // Wk:Kw *x̣ím'a* 'buttocks' // NEC **χħwɨ̆mχV* 'cheek, buttock'.

(9) COLLAR-BONE: IS:Sp *(s-č'im-)aɬ-q'ʷl-t* (root *q'ʷl-*) // NEC **qHwŏɬwV* 'neck, collar' (> AvA **qʷilu*).

(10) (?) HEAD: NWk:Kw *xum-s* NEC *hq'wĕmV̄* 'horn, head'.

(11) BODY: IS:CO *=ic'aʔ* // NEC **čōrχV*

(12) FAT: NWk:Kw _cnx̣_^{W}_-iʔ_, noun > Kw _cénx̣_^{W}_aʔi_ (noun) // NEC (And.-Darg.) _*c'ēnx̱wV_ (also _*c'ēnλV_), adj.

(13) HEAD: CS:Se _c'əq'_ (root; from _*c'iq'-_ ?) // NWC _*sq/'a_ / Yen _*ciGV_ (_*c < *c'_) [Cf. NWk:Kw _səq'a_ 'above'; cf. also Ath _-ciʔ_, etc.].

(14) MOUTH: NWk:Kw _sem-s_ // NEC _*ʒ́wĕmV_].

(15) ARM, SHOULDER: CS:Ls _dəx̣_ 'arm' (_d < n_) : Sq _nix̣_ id. // NWk _n'ik_^{W}_-_ in He _n'ik_^{W}_-lá_ 'carry on the shoulder' // NC _*nHɨ̆wG̱Ā_.

(16) FOREHEAD: Ts:UC _ƛ'ox_^{W}_-s_ // NEC (AvA and Tsez.) _*ƛ'arq'wĕ_ > AvA _*ƛ̣aq̇_^{W}_ara_ (> Av. _ṭaɣúr_) / Tsez. _*ƛ̣ɔq̇ɔ_ forehad, cap'. [Sa _-x_^{W}_-_ may match Av. _-ɣu-_; Sa _*ƛ' < *ƛ'_; Sa _o_ < _*a_].

[=16] UC _ƛ'ox_^{W}_-_ 'fhd'	Av. _ṭaɣúr < *ƛaq̇_^{W}_ara_	Tsez. _*ƛ̣ɔq̇ɔ_ 'forehead'	< NEC _*ƛ'arq'wĕ_ id.

(17) HIDE: IS:Th _ɬex_^{W} 'patch' : CO _ɬx_^{W} 'dress' : Sp _-eɬx_^{W} : MS _-ɬx_^{W} 'skin, clothes' // NEC _ɬĕʔ/ʔwnɨ-_ 'skin (of an animal)'.

(18) EAR: IS:MS _leš-n'_ 'hear' (<_*lex_): Ka _leš-ən_ 'hear' / CS:Ls _luh_ 'hear' // NC _*ɬĕHɬe/i_ or _*ɬĕHe._

(19) FAT, GREASE: IS:Th _miʕ'_^{W}_-_ 'grease' // NEC _*măɦwV._

(20) HEAD, FACE: Ts:UC _matin_ 'head' // NEC _*mət'e/i_ 'face' [// N _*mEt(')a_ 'head, top'].

(21) HORN: Ts:UC _winaw'_ (_i_ may originate from _*i/e/ĕ/ē_); cf. CS:Se _wenaʔəw_ // NWk:Kw _weƛ'ax̣_ 'horn, antler' // NEC _*wĕnƛ'V_ 'beak, horn; head'. [_*-nƛ'-_ > Wk _-ƛ'-_ : Sa _-n(V)ʔ-_ ?].

VII. Shift of [w] in Pre-Salish When Compared with NC

Sa often shows labials _q_^{W}, _q'_^{W}, _x̣_^{W}, _ʔ_^{W} where NC has no [w]. Appropriate NC words contain _*Cw_, _*Lw_. A pre-Sa methathesis may have simplified the pronunciation: it was easier to use [w] as a part of _q_^{W}, _q'_^{W}, _x̣_^{W}, _ʔ_^{W} than as a part of _*C_^{W}, _*L_^{W}, so [w] shifted from [C, L] to [q, x̣].

(1) BEND, CURVE: CS:Se _k'_^{W}_uc'-_ 'bend' (v.) // NEC _*kɨ̆ɛ̆'wɨ̆_ (also _*č'ūkwV_) 'curved'.

(2) SHARP: CS:Ls _x̣_^{W}_əc_ // NC _*ɦüǯwĀ_ > Tsez. _*ʔãč-_; Lezg. _*ʕeč:_^{W}_a-._

(3) FLOW, POUR, WASH: IS:MC _c'aw'_ 'wash' : Sp _c'ew'_, _caw_ 'wash' / Ts:UC and CS:Se _c'əx̣_^{W}_-_ 'wash' // NWk _cx_^{W}_-_ in Kw _cx_^{W}_la_ 'overflowing' etc. // NC lex. suff. _*=Hăć̱wV_ 'pour, wash' (> _*ć̱wăHA_ ?). [Note that _*w_ shifts from _*C'_ to _x̣_ in Sa, and to _x_ in Wk].

(4) (?) POLE, STICK: NWk _c'wax̣-_, Kw _c'ux_^{W}_-_ 'insert (pole)', _c'x_^{W}_-_ 'stab' (cf. NWk _zux_^{W}_-_ 'log, pole') // NEC _*c̱'wēχV_ 'stick, chip, piece of wood, beam''.

(5) SHOULDER: NWk _n'ik_^{W}_-_ 'carry on the sh.' // NC _*nHɨ̆wG̱Ā._

(6) BERRY: NWk *n'ux*^{w}- (Kw *n'ux*^{w}*a* 'small blueberry') // NEC **niwG̱V.*

(7) ROCK: IS:Th *ƛ'ix̣*^{w} 'rock, gravel' // NEC *ƛ̱'wĕhrŭ* (/-*l*-) 'rock, cliff' [*w*-shift; Sa -*x̣*^{w} : **-(h)rŭ*].

(8) HARD: Ts:UC *ƛ'əʕ*^{w} (:*x̣aƛ'e·*- id.) / IS:MC *yəʕ*$^{\prime w}$ (note shift of glottalization *ƛ'* to *ʕ'*; note *ƛ'* : *y*; *yəʕ*^{w}, *y* < **L*) / Ts:UC *ƛ'əx̣*^{w} / CS:Sq *ƛ'əx̣*^{w} / BC *ƛ'ax̣*^{w} // NC. **Ƚ̱wĕrV.* [**w* shifts from left to right, after **r* turns to *x̣*]; MC *ʕ* may originate from **r*, and *y* from **L*].

VIII. Sporadic Spirantization of the Underlying Uvulars

**Gw*, **GHw*, **q'w*, **q'Hw* in Salish

Spirantization of uvulars, especially labiouvulars, is rather widespread both in NC and Sa, but this process is much more frequent when the voiced uvular **Gw* is involved:

(1) IS:Th *q'*^{w}*ət-x̣*^{w}*əm'* 'joint' vs Ts:UC *x̣*^{w}*ut'* 'bent up' // Tsez. **q̇ɔ(n)tV* < NEC **q'HwəntV* 'knee/elbow'. [Note unstable pre-cons. nasal in NEC languages vs possible loss of **n* in pre-Sa].

(2) IS:MC *q'*^{w}*ic'*- 'twist' vs. IS:Sh, CS:Sq *x̣*^{w}*əc'-x̣*^{w}*əc'* 'joints' : CS:Se *x̣*^{w}*əc'-q'*^{w}*-uya* 'wrist' / Ts:UC *x̣*^{w}*uc'* 'bent up' // NEC:AvA-Lezg. **q'HwĕmčV* 'hook, curved' with unstable **-m-* in Lezg. *q̇/*^{w}*a(m)č̣V* 'bend, elbow, tip, point'. [NEC **q'-* doesn't undergo spirantization; cf. also next ex.].

(3) STICK: IS:Th *x̣əc*, *x̣ic* 'wooden' / CS:Ls *s-x̣ac* // Tsez *ħiš* < TsKh **χ₁ešu*, **χ₁öšu* < NEC **GHwälćV* 'stick, board; bolt'. [Spirantization and delabialization both in Sa and Tsez.].

(4) DEER: IS:MC *x̣*^{w}*əl(')*- 'buck' / CS:Ls *x̣*^{w}*el* 'deer' // Gin. *ʁ*^{w}*il* 'doe' < Tsez. **ʁ:*^{w}*el* id. < NEC **Gwăłā* 'doe, hornless goat' (> AvA **ʁ*^{w}*alV* 'hornless goat/ram'). [Spirantization both in Sa and in NC daughter languages (Nakh., AvA, Tsez.)].

(5) (?) ASK: Ts:UC *ʔo·x̣*^{w}*aš* (*š* < *x*) // NEC (AvA-Lezg.) **HreqwȀ(r)* > Lezg. **ʔerχ*^{w}*a.*

(6) EAGLE: Ts:UC *ƛ'ix̣in* / CS:Ls *yəx̣*^{w}*(-)əlaʔ* [NB *y* < *ƛ*] // Lezgi *leq̇*, Ag. *liq̇/* < Lezg. **lïq̇/*^{w} < NC **lHĩq̱'wȀ* 'eagle', etc. > NWC **la(r)q*^{w}*a* (> Abaz. *laħ*^{w}*a*, Ub. *daχ*^{w}*â*). [Note delabialization in Ts:UC vs NEC:Lezgi-Ag., and spirantization *x̣*^{w} in CS:Ls vs *ħ*^{w} in Abaz. (and Abkh.), and *χ*^{w} in Ubykh]:

UC *ƛ'ix̣*-,Ls *yəx̣*^{w}	Lzg. *leq̇* < **lïq̇/*^{w}	A *laħ*^{w}*a* U *daχ*^{w}*á*	< WC **la(r)q*^{w}*a*	< NC **lHĩq̱'wȀ*

(7) SEW: IS:MC *łəx̣*^{w}- : MS *ɬx̣*^{w} // NC **=ilq̱'wVn* 'stick into, sew'.

We may compare a case (ex. 8) where Sa -*w*- matches NC *-*Gw*-, indicating a shift **Gw* > **ɣw* > *w* (cf shifts of the type *x*^{w} > *w* in Sa languages):

(8) IS:Th *maw-e* 'gossip' (< Sa **maɣw-* ?) // NC **maGwV* 'word, sound, song'. - Cf. Sa *ɣ* < **(r)G*, ex. 9. - Altern.: Th *mawe* : NEC **mħārχwā* 'tale' > Lezg. **max*^{w}, etc.

(9) IS:Th *məyew'* 'lynx' // NEC **mHarGVwV* 'tom-cat' (> Lezg. **marq:/aw* > Ag. *marRu*). - Cf. NWk:He *mauẋwa* 'bob-cat'.

IX. Reduction of Underlying Clusters in Salish:

-VC < **-VRC-* (**R* = *n/m*, *l*, *r*)

On many occasions, Sa shows reduction of prehistoric intervocalic clusters to single consonants; the cluster itself may be a reduced form of an underlying sequence *-CVC-*:

(1) [*t* < **nt*] IS:Th *q'*^{W}*ət-x̣*^{W}*əm'* 'joint' (cf. Ts:UC *x̣*^{W}*ut'-* 'bent up') // NC **q'HwəntV* 'knee/elbow'.

(2) [*k* < **nk*] GOOD: NWk:Kw *ʔik* // NEC **ĭnk̠wV* 'right, good'.

(3) [*c'* < **mĉ*] MOUNTAIN GOAT: NWk:Ha *c'aG* (< **Gac'* ?) // NEC (AvA-Darg.) **Gamḉ̠ā* > Darg. **q:ača* (with *č:* in some dialects).

(4) [*c* < **Nc* < **mVc*] ANT: BC *qac-qλ̵* // NEC:Nakh. **qēç̣ʒ-* AvA **ʁamç̣a* < NC **q/GāmVc'V* (also 'grasshopper'). [Cf., in North America: Tsimshian *s-gāns-gōʒinx̄t* 'ant'. - Note also N **K'[u]č'V* 'ant']

(5) [*l* < **mλ̵*] MOUNTAIN GOAT: NWk:Kw *p'lx̣* (root) // NEC **bHēmλ̵̠i.*

(6) [*c'* < **rć*] ARM, EMBRACE (etc.): CS:Sq *q'ac'* // NC **qărć'wV* 'shoulder, arm'

(7) [*c* < **lć*] STICK: IS:Th *x̣əc, x̣ic* 'wooden' / CS:Ls *s-x̣ac* // NC **GHwälćV* 'stick, board'

(8) [*s* < **ls*] TO SCRATCH: IS:Th and MC *qəs-* // NC **qāls̠V* (> AvA. **qas:-*, Darg. **qars:-*)

(9) [*ɬ/r* < **rH*] COLD: IS:MC *c'əɬ-* : Sp *c'er* // NC **č'wErHV.*

(10) [*q*^{W} <**rq'w*] D̦IRT: NWk:He *miq*^{w}*a* 'dirty, muddy' / NC *mHĩrq̠'wV* 'dirt, rust' (Lezg.:Kryz *meq*).

X. A Few More Examples of Simplification of the Underlying Roots in Salishan

Table I shows nine cases of simplification of original clusters in Sa: Underlying **c'w-/*c'ɦw-/*Hc'w-* becomes *c'-* (and *c-* in a compound) in Sa; **ɦwm-* becomes Sa *m-*; **q'Hw-* becomes Sa *q'*^{W}-, etc.

TABLE I

a EC **Hc̠' wēj mə̆* leg-bone PSa **(s-) c' u̦/a m'* bone	b EC **-c'ɦ wē me* eyebrow CS:Se *-c u m-* eyebrow	c NC **c' ā̃ nHV* arrow No *c' i·* *ḥ-* arrow
d EC **c' wĭ ɬɦV* stick, branch IS:MC *c' ə l,* MS *c' i l* tree	e EC **ɦwm ă ɦwā* moist,pool IS:Sp *m o ʕ*^{W} flow	f EC **q'Hw ĕ mç̌V* curved IS:MC *q'*^{W} *i c'-* twist

g EC **q'ɦwĕɬV̄* cow, mare IS:MS *q'*^{w}*ey-q'*^{w}*ay* bison, buff.	h EC **ƛ̱HwemV* liquid (adj.) CS:Sq *ɬəm?x*^{W} rain, *ləm?* dew	i EC **mh ă lV* warm (/-*n*-) IS:CO *m*ʕ *a l* (:*ma*ʕ*l*) warm

[To ex. a: Cf. Eyak-Ath **c'em?* 'bone' [] b: EC root is preceded in this compound by **c'ilV-* 'eye' (:CS:Se *c'il-* 'look for') [] c: Altern.: NEC **ć'ăʕV* 'reed, cane; arrow' [] d: Cf. Wk:No *c'aɬ-aq* 'branching out' [] g: Cf. Th *q*^{W}*is-p* [devoicing] [] h: N **LaHm u/* 'marsh, silt, wet' (St. '89#106) [] i: Cf. NEC:Ts. *mɔʁ:V*]

Cf. some other examples of simplification of the underlying roots in Sa (end of the underlying roots seems to be lost):

(1) EAR, HEAR: IS:MS *leš-n'* 'hear' (<**lex*): Kalispel *leš-ən* 'hear' / CS:Ls *luh* 'hear' (cf. Ts:UC *s-lix-n* 'ear-lobe') // NC **ɬĕHɬe/i* (**ɬ* because of Hurr. *lēlə* 'ear') or **ɬĕHe* (if Hurr. *-lə* is a suff.).

(2) HIDE (Sa has both a regular word and a lex. suffix): IS:Th *ɬex*^{W} 'patch' : CO *ɬx*^{W} 'dress' : Sp *=eɬx*^{W} : MS *=ɬx*^{W} 'skin, clothes' // Lezg. **le?*

'skin' < NEC *ɬĕ?/ʔwnɨ* 'skin of animal'.

(3) SWALLOW, THROAT (etc.): IS:Th *məq'* 'satiate', *məq*^{W} 'hold in mouth' (:CS:Ls *bəq'* 'put/hold in mouth, swallow'; *b* < *m*) // NC **mVq'VɬV* 'throat, larynx'

XI. Some Salishan *CVCVC*-type Roots and their NC Cognates

In many cases Sa roots of the type *CVCVC* seem to have developed the part -*CVC* from underlying clusters -*CC*-. Some roots show late suffix-like additions.

Example 1 seems to show -*CVC* from -*CC*-, as well as a reduction **Gw* > *q*, possibly, in an unstressed position. Sequence *-x̣*^{W}*aq'*^{W} (secondary labialization in *x̣*?) in example 3 seems to match NC **-rq'w-*. Sequence *-qiw* in example 4 may relate to NC **-gw-*; cf. *-saw-* vs **-św-*; in example 5 (note here also *y* vs **r*). For the process -*CVC* < -*CC*- in pre-Sa, cf. AvA:Lak. *ç̣iniq* < NEC **c'ānq'V* 'lynx, panther' (etc.).

(1) ANGER (etc.): Ts:UC *qəɬ(ə)x̣-* // NEC **Gwāɬħo/ə* 'gossip, offence, anger' (UC *x̣* : NEC **ħ*).

(2) (?) WAR: CS:Ls *x̣ilix̣* // NEC **Ƚ̱ĕɬV̆*. [Sa may show *x̣* vs NC *L*-type cons.; same in NC].

(3) (?) ELK, GOAT (etc.): IS:Th *təx̣*^{W}*aq'*^{W}- (first stem in 'doe') // NEC **dVrq'wV* 'he-goat'.

(4) HORSE, DONKEY: Ts:UC *tiqiw* 'h.' / CS:Sq *s-taqiw* 'h.' // NEC **dhogwā* 'd.' < (?) NC **t'HōgwV* 'hoofed animal' (covering also NC **t'ūgV̄* 'he-goat').

(5) TREE: Cs:Ls *yesawi* 'alder' // NC **räśwē* 'tree, wood'.

In examples 2 and 6, *auslauting* consonants may represent a relatively late addition. Sa word in ex. 2 (*x̣ilix̣*) seems to show structural and phonetic symmetry, typical for Yokuts (Penutian).

(6) EAGLE: / Ts:UC **ƛ'ix̣in* (?) (:CS:Ls *yəx̣^{w}əlaʔ*) // NC **lHīq̱'wǍ* (also with **ɬH-*, **-q̱-*) > Lezg. **liq'l^{w}* > Lezgi *leq'* / Darg. **ʔiχl-liq'an*, loss of labialization, as in Ts:UC?.

(7) LAND, EARTH: CS:Ls *ɬalil* 'to go ashore, to land' // NC **lhĕmŁ̠wɨ̌* 'earth'. [The underlying cluster **-mŁ̠w-* is probably simplified to *-l-* in Sa; *CVCVC*-status is acquired by adding *-(V)l*].

Example 8 shows *CVCVC* becoming a *CVCC*-root, but it still behaves as a two-vowel root, namely, *CVC[V]C*, - otherwise it would lose the nasal (**-VmC-* > *-VC-*). - Example 9 is usually interpreted as root-type *CCVC*; it is still pronounced as *CVCVC*.

(8) CS:Ls *ɬəbc'-* 'mucus' (*b* < *m*) // NEC **ƛămVč'V* (*/x-*) 'saliva, pus' > AvA **ƛa(m)ç:V* 'saliva'

(9) IS:Th *məγew'* 'lynx' // NC **mHarGVwV* 'tom-cat' [Cf. Wk:He *mauх̣wa* 'bob-cat'].

XII. Salishan Roots that Match N(E)C Lexical Suffixes

N(E)C lexical suffixes frequently match both Sa roots and Sa lexical suffixes. NC lexical suffuxes show genetic links to Sa much more frequently than "regular" N(E)C roots do. N(E)C lexical suffixes may correspond either to lexical suffixes or to regular roots in N(E)C daughter languages.

We may deal with pre-Sa metathesis in examples 1-4.

(1) [Type *C^{W}Vy* in Sa vs *=iCwV* in NEC] GO: IS:MS *x^{W}u, x^{W}uy* 'go' // NEC lex. suff. **=iχwV*

(2) [Type *C^{W}Vw* in Sa vs *=VCwV* (*-V* = *-U* ?)] BURN: Ts:UC *kəw*, *k^{W}əw* // NEC:East Dag. **=ōgwV* (> Lezg. **ʔok:^{W}ɨ* / Khin. *=ek:-* / *k:-*).

(3) [Type *C^{W}VC(V)* in Sa vs *=iCCwV* in NC (> *=iCVC* ?)] RIPE(N): Ts:UC *q'^{W}əli* 'ripe' (<**ilq'^{W}ə*, with *l* from **r* ?) // NC **=ĭrq'wǍ* 'ripen' > NEC:AvA **=iq̇(-Vl)-* 'ripe(n)'. [Cf. frequent *l* < **r* in Sa; pre-Sa **q'wirV* ?]

(4) [Type *CVw'* / *CVH^{W}* in Sa vs *(? CwVHV <) =HVCwV* in NC (> *CVw*)] POUR, WASH: IS:MC *c'aw'* : Sp *c'ew'* 'wash' / CS:Se *c'əx̣^{W}* / Ts:UC *c'əx̣^{W}* [from underlying **c̱'wäHA*?] // NC **=Häc̱wǍ* 'pour, wash', or **=Häc̱Ǎw* > AvA **č:Vb-*. [**w* shifts from *c'-* to *x̣-*. - Note that both Sa and AvA *CVC* matches NC root type *=HVCVw* (rather than *=HVCwV*)].

(5) [Type *CVC^{W}* in Sa vs *=iCCwV* in NC] DIE, KILL: IS:Th *ƛ'əx^{W}* 'cripple' : MC *ƛ'əx^{W}-* 'die, kill' / (?) CS:Ls *ɬəx̣^{W}-* 'hunt' (also IS:MS *ɬx^{W}-*) // NEC **=ilχwV-* (*/-qw-*) 'die' vs **=ilχwV-* 'kill'. [Note sound symbolism in NEC].

[=5] Sa:MC *ƛ'əx^{W}-* 'die, kill' / ? Ls *ɬəx̣^{W}-* 'hunt'	NEC **=ilχwV-* (*/-qw-*) 'die'; **=ilχwV-* 'kill'

(6) [Type =VCwV in NEC] WHITE, BLUE: Ts:UC *q^{W}a/ux̣^{W}-* 'wh.', *q'^{W}ix̣-* 'b.' // NEC **=ĕq̱wǍ* 'yellow' > AvA **q:oʔi-* 'wh., b.' >Av. *q:áħa-b* 'wh.', Chad. *qáħa-b* 'wh.', And. *q:^{W}oj* 'b.' (PAnd. **ʔV-q:^{W}o-ji-*).

(7) [Type *C*^{W}*VC(u)* in Sa vs *=HV-CwVC* in NEC] FEAR: Ts:UC *q*^{W}*anu-* // NEC (AvA and Lezg.) **=Hä-GwVn* 'tremble, be afraid' > AvA **=iʁVn- / *ʁiʁVn-* > Cham. *ʁiʁin-*, Tind. *ʁiʁan-* / Lezg. *ʔäq:/*^{W}*Vn-* > Arch. *e=q/*^{W}*in-*.

(8) [Type *CVC*^{W} in Sa vs *=iCCwVC* in NC; altern.: Sa *CVC*^{W} is a metath. of *CwVC*] SEW: IS:MC *ləx̣*^{W}*-* : MS *ł̣x*^{W} // NC **=ilq̇wVn.* - Pre-Sa may represent a metath. of the old **q'wil-*.

(9) [Type *wVC* in Sa vs *=VwCwV* in NEC] TO OPEN: Ts:UC *wał-a-* // NEC **=ĕwλ(w)V* > Lezg. **ʔäwλɨ* 'unlock'. - (Cf. Ts:UC *wax̣-* 'open (eyes)' // NEC **=ĕwλ(w)V* > Darg. **ʔawx:-* > *ʔabx:-* 'to open').

(10) [Type *CVC*^{W} in Sa vs *=iCCwV(C)* in NC] BEND: IS:Th *łaq'*^{W}*əw-t* 'bend over' // NC **=ilq̱'wV(ł).*

(11) [Type *CV* in Sa vs *=VCwV(C)* in NC] DO, MAKE: Ts:UC *wi* // NC **=ăhwV(r)* [ablaut **a/i*] > NEC:Tsez. **=Vw-* / NWC **wə.*

[=11]UC *wi-* do, make	Tsez. **=Vw* id.	WC **wə* id. <	NC **=ăhwV(r)* id.

(12) [Type *CVC(V)* in Sa vs *=VCVC* in NEC (> *CVC*)] SUCKLE, CHEW: Ts:UC *c'am-i-* 'suckle' // NEC **=ĕč̣Vm-* 'gnaw, chew' >AvA **č̣Vm-* 'chew' / Darg. **č̣a/m-.*

(13) [Type *CV* in Sa vs *=VCCV* in NC (> *VCVC*)] SPEAK, TELL, TALK: Ts:UC *cu-t*, *cu-n* (<PSa) // NC **=[ĭ]mcŪ* > AvA **=ocVn-* / WC **c:*^{W}*a*

[=13] UC *cu(-n)* <PSa	AvA **=ocVn-* 'talk'	WC **c:*^{W}*a* 'talk'	< NC **=[ĭ]mcŪ* 'talk'

(14) EYE, LOOK: Ts:UC *=us-i-* 'face, eye', *-al-us* 'eye' // NEC **=HimsV* 'to look' (> AvA **-us-* > *-us-* 'to search' in modern AvA l-ges) [cf. NEC **c'il-* 'eye' : Sa **c'il-*]

[=14] Sq *=us* 'eyes',UC *=us(-i)-*	*-us-* < AvA **-us-* 'to search'	< EC **=HimsV* 'to look'

(15) [Type *yVC* in Sa vs *=iCCwV* in NC] (TURN) AROUND: IS:Sp *yir* (and a borrowed *yal* ?) / Ts:UC *yəl* (*l* < **r*) 'around' // NEC **=irʁwV* 'roll, turn around'.

(16) [Type *yVC* in Sa vs *=iCVC* in NC] SPIN, WEAVE: CS:Ls *yiq'-* 'weave, knit, spin' / Ts:UC *yəq*^{W}*-* 'twist, spin' // NC **=iqăr* 'weave' (verb preserved only in WC and E.Daghestan languages) > NEC **wiqVrHV* > **qwVrHV* '(smthg) woven' > Darg. **q*^{W}*a/r* > Chir. *qu/r* 'horse-cloth' (etc.).

(17) [Type *yVC* in Sa vs *=iCwV* in NC] RETURN: Ts:UC *yac'-* 'turn back, turn around and come back' // NC **=ić̣wĔ* 'come, return'.

(18) [Type *ʔiC(C)* in Sa vs *=iCwVC* in NEC] EAT: Ts:UC *ʔiłn* // NEC **=iʡwVl* 'feed on, eat' (> Tsez. **he n- / *he/l-*).

(19) [Type *ʔVC*^{W} in Sa vs *=iCwV* in NEC (> *=VCwV*)] GO: CS:Ls *ʔux̣*^{W} // NEC **=iχ̣wV* / **=iq̱wV* 'go, come, enter' (> Tsez. **=uχ:*^{W}).

(20) [Type *ʔVC* in Sa vs *HVCCV(C)* in NEC (> *=VC*)] SEE: Ts:UC *ʔəx̣, ʔa-ʔx̣n* // NEC **=Hārq̱'V(n)* 'see, find' (> Tsez. **=ɨ̃q̣-* / Darg. **=ah/-*).

(21) [Type *CVC*^{W} (with redupl.?) vs *=VCV* in NEC (> *CVCV* with redupl.)] QUICK: Ts:UC *x̣ax*^{W} 'quick, hurry' // NEC **=ă̈χV* 'quick, swift' (>AvA **χ:iχ:V-* > Cham. *χ:iχ:u* id., Cham. Gig. *χ:eχ:a* 'quickly'; similar in other AvA languages). [Sa almost equals AvA].

[=21] UC *x̣ax*^{W} 'quick'	Cham. *χ:iχ:u* 'quickly'	< AvA **χ:Iχ:V-* <	EC **=ă̈χV* 'quick'

XIII. NC Words with **r* and Their Cognates in Sa Languages

One of a few significant changes from PSa to individual Sa languages is the evolution of the underlying **r* and its variants. Even in conservative Sa languages, the reflexes differ, being realized as *r, l, ɬ, ƛ', y,* (in many cases:) *x̣, x,* ʕ, *ʔ*, ; consonants *x̣, x,* ʕ, *ʔ* may be labialized ([w] represents a neighboring labial vowel or an underlying labial consonant).

(1) DIRT: Ts:UC *q'ax̣-* 'mud, smear, paint' // NC **q'ă̈rē* (**r* > *x̣* in Sa?).

(2) BANK: MC *xər'-xər'-t* 'steep bank' (root metath.?) // NEC **rēGV.*

(3) DAY: Ls *ləx̣* // NC **rĭhV.* [Sa *l* may originate from **r*].

(4) HARD: PSa **ƛ'a*ʕ^{W}*/x̣*^{W}*-* > Ts:UC / CS:Sq *ƛ'əx̣*^{W} / BC (Nu) *ƛ'ax̣*^{W}*ƛ'əx̣*^{W} vs IS:MC *yə*ʕ$^{/w}$ // NC **Ł̱wĕrV* ([w] shifts to the right, to a more "comfortable position", after **r* turns [x̣]; there is also a shift of glottalization from *L' - H* to *Y - H'*).

(5) WOOD: CS:Ls *yesawi* 'alder' // NC **rắśwē* 'tree, wood' (Sa *y-* < **r*, **-Sw-* becomes *-saw-*). [(6) BREAK: IS:Th *ma*ʕ' : CO *m*ʕ$^{(w)}$: MC *ma*ʕ$^{/w}$: Sp *maw'* // SC:ST **muar* 'bite' vs N **murV*, St. '89 #127 (shift of the type **mwar* > **marw* in pre-Sa?)].

[(7) (BE) VERY SICK, DIE: IS:Sp *mix̣*^{W} : MS *mix*^{W} / CS:Ls *mix̣*^{W} (< [mIx̣/rU]?) // N **märV*]

There are many other Sa roots that can be linked to Kartv., some of them may indicate old SC borrowings to Kartv., but most forms seem to reflect genetically related, inherited roots.

The following examples cover the shift from an underlying cluster of the type **X(w)r* to Sa (labio)uvular sounds, velar sounds (including *ɣ*), *ʔ* (> '); relatively seldom to *r - ɬ - y.* (This is similar to the development of **r* in Sa)

[(8) OLD: IS **kix* 'close elder fem. relative' // SC:ST **Kri* // N **kirHA* (St. '89 #50)].

(9) HORN: (I) IS:Th *q*^{W}*ay'* (*y* may originate from sonorants of the type *y, l, ɬ, λ', r*, *-'* in *-y'* may indicate an underlying **H*) / (II): IS:MC *qəx̣-*, *qx̣-* (+ stressed "instrum." suff. *-min*; loss of [w] in an unstressed root in a compound *C(V)C-CVC*) : Sp and MS *qx̣-* (with suff.) // NC **qwĭrhV* (Sa stem I **q*^{W}*a/ir'* seems to match NC) // N **kErV* (NB Sa stem II); St. '89 # 86.

(10) COLD: IS:MC *c'əɬ* 'cool off' : Sp *c'er, c'al*; cf. Ts:UC *c'ix̣*, etc. // NC **č'wErHV.*

(11) BOY: IS: CO *t-twi-t* : MC *tw'i-t* : MS *t-t'wi-t* : Th *təwi-t* (dimin. *tʔu-t* 'little boy' may come from a Sa root for 'little') / CS:Ls *tawix*^{W}*-* 'child/offspring' (root in pl. form 'children') // NEC **dwirχE* 'boy, son'. - As for IS:MC *t'a*ʕ$^{/w}$*-* 'little', Th *tʔu-* (etc.), cf. NC **t'iHV* 'small, little' (possibly = NC **t'iHU* > AvA **t'VH*^{W}*V-*).

(12) (?) GROW(TH): PSa **ƛ'aẋ*^{w} (as both in IS and CS) 'grow(th), old' > CS:Ls **ƛ'aẋ*^{w} 'grow(th)' - Also IS:CO *ƛ'ẋ* : MC *ƛ'əẋ-* (root) 'grow up' / CS:Se *ƛ'aẋ-aẋ* *'grown up' > 'old person' // NEC **ƛ'ŏrẋwV* 'sprout'. (Sa **a* is shown by both main branches: IS:CO and CS:Ls).

(13) ROCK: IS:Th *ƛ'iẋ*^{W} 'rock, gravel' // NEC *ƛ̱'wĕhru* (/-*l*-) 'rock, cliff' [Note **w*-shift, from left to right in pre-Sa].

(14) (?) INTESTINE: MC *p'i?-p'i?* 'guts, int-s' // NEC **bħĕrƛ̱'V* 'large intestine'.

(15) LYNX (etc.): IS:Th *m(ə)γəw* // NEC *mHarGVwV* 'tom-cat'.

(16) TELL A STORY: IS:CO *m'ay?-*, *m'ay'a?-* : *malẋa?* 'lie' (a different root?) / MC *may'-* (root) 'tell a story, confess' // NEC **mˤ/ħārẋwā* 'tale'. [CO *m'-* seems to match NC **mˤ/ħ-*].

Conclusion

The Salishan(-Wakashan) languages seem to originate from a prehistoric language (or languages) that was (or were) very similar to North Caucasian (being later "torn away" as a result of some prehistoric migration, ending up in northwestern America?). Salishan languages may or may not be a part of the North Caucasian language family [as represented by NCED]; on occasion they show close parallels either to Avar-Andian, or to Lezgian, or to Northwest Caucasian languages: for instance, Sa:Sq *ɬəm(?)* 'dew' (cf. *ɬəm?-x*^{W} 'rain') matches easily AvA **λ:imV-* 'liquid' (as in And. *λ:emi*) and Lezg. **λ:älmä-* (but not Lak. and Darg.), as well as the more complex proto-form NEC **ƛ̱HwemV* 'liquid' (adj.); this cannot be a coincidence, there are too many such precise correspondences. (cf. also numerous Wk-NC exact matches, such as He *q'*^{W}*en-G*^{W}*a-* 'throat' vs AvA **q̇:*^{W}*an-q̇:*^{W}*a* < NEC **Gwan-G̱wa* 'throat'; or NWk = Kw *cnẋ*^{W}*-* 'fat' vs NEC = AvA-Darg. **c'ēn̲xwA* 'fat'; or No *k*^{W}*itḥ-* 'tap, knock' vs AvA **k*^{W}*irṭa* 'hammer' [> Tind. *koṭa*, etc.] / NWC **k:əṭ*^{W}*V* 'axe').

We would expect such a state in a separate group(s) of languages that are closely related to the languages in question (i.e., NC). Lexico-grammatical material of Salishan languages – which connects these languages with NC – is enormous.

Besides the Salishan-Wakashan languages there are several other languages in North America that are related to the North Caucasian (or, more broadly) Sino-Caucasian languages of Eurasia: first of all, some language isolates: Chemakum, Kutenai, etc.; Eyak-Athapaskan, Tlingit, Haida, Algic (and possibly also Keresan and Siouan), and others.

Toward Reconstructing Proto-South Khoisan (PSAK)

Christopher Ehret
University of California/Los Angeles

My goal in this article is to accomplish two things. The first is to report on a work in progress, a comparative historical reconstruction and etymological dictionary of South Khoisan (Southern African Khoisan), and on the preliminary sound correspondences and sound shift rules that can be formulated from that data. The second is to raise some very strong points on the full applicability of the comparative method to Khoisan and on how well it appears from the available evidence to work for the Khoisan family.

A work in progress

The year before last I undertook the task of building up a systematic phonological reconstruction of the proto-South (proto-Southern African) Khoisan (SAK), drawing on the materials available in new first-rate dictionaries of key languages from the Zhu and Taa-!Ui branches of Southern African Khoisan family (Dickens 1994; Traill 1994), as well as on the insights and new discoveries of Rainer Vossen's reconstruction of the Khoe branch (Vossen 1997). In the spring of 2002, I put these materials together in the form of a preliminary draft etymological dictionary of 1500-plus proposed roots, in order to present them for discussion at a workshop at the Santa Fe Institute held in the summer.

Late in 2002 I went back through the draft dictionary, correcting mistakes, removing redundancies and unsustainable cases, and adding a number of further etymologies. The draft dictionary as it now stands also includes two kinds of supplemental information, which I have added at hundreds of points in the dictionary:

1. At the summer workshop meeting, I had presented proposed sets of old verb and noun suffixes in Khoisan. Some remain productive in modern-day languages. Others are no longer productive but can be postulated through either internal or comparative reconstruction. Wherever the reflexes of these suffixes appear to be present in the proposed cognate of the etymological dictionary, a notation to this effect has been added to the relevant entry.
2. In addition, I have entered semantic explanations in the etymological dictionary wherever the semantic connection did not seem transparent.

This is not the time or place to present the currently 325-page-long etymological dictionary. I am hoping to turn it into a completed monograph with text by the end of 2003. For the present, however, we can preview something of the directions of research the etymological dictionary permits us to pursue.

The comparative data contained in the dictionary allow us to propose rigorously regular sound correspondences among the well-studied languages—in the segmentary phonology to a virtually neogrammarian standard—despite the great time depth that separates the branches of SAK from each other. The most probable archaeological correlation of the initial spread of SAK languages with the Wilton archaeological horizon places this time depth at around 8,000-9,000 years (Ehret 1997, 2000).

There are strong internal indications of the overall validity of the scheme of SAK phonological reconstruction:

a. The consonant correspondences as revealed in the etymological dictionary allow one to construct general sound shift laws, along natural directions of sound change, that explain the consonant outcomes in systemic, ordered, and parsimonious ways.
b. The vowel sound correspondences normally follow fully regular patterns as well, although with sporadically free-variant outcomes for the set of reconstructed front-vowel-final diphthongs.

One pair of vowel outcomes remains to be fully explained at this point: Usually PSAK *o > !Xoõ (Taa-!Ui) u /C_C[+labial]. But in a significant minority of cases involving this particular environment, PSAK *o remains /o/ in !Xoõ. The governing environments of these alternative outcomes apparently need further specification.

The proto-South Khoisan (PSAK) language was surely tonal, just as are its modern-day descendants. With great tentativeness one can propose a system of four tones in roots of two or more morae. Each mora of a word in this system carried a tone. For now, five tone distinctions are needed to account for tonal correspondences in single-mora roots, but we may well eventually be able to identify conditioning factors allowing us to simplify the tones to four in those roots, too.

The patterns as revealed in the etymological dictionary leave one big problem as yet unresolved, namely the correspondences among the secondary articulatory features of vowels in early SAK: nasal, pharyngeal, breathy, laryngeal, and sphincteric. One proposal I will be investigating is that the instances of non-regularly correspondng secondary features in stem vowels originated as features belonging to the vowels of suffixal morphemes added to the stems, with these secondary features of the suffixal vowels then spreading leftward to the stem vowels. This proposition would make phonetic sense because it is characteristic of the secondary features in SAK languages to occur right across the vocalic morae of the word.

Consonant sound correspondences and sound shifts

PSAK had very complex array of consonant articulations, including numerous word-initial, complexly articulated consonants as well as numerous word-initial consonant clusters. We will first lay out the systematic sound correspondence patterns among the three well known SAK languages—!Xoõ for the Taa-!Ui branch, Ju-/Hoan (often shortened to "Ju" here) for the Zhu branch, and Nama for the Khoe branch. Subsequently we will go on to propose sound shift histories and to consider the question of which of the multiply articulated items might be single consonants and which might be clusters.

Table 1 lays out click consonant correspondences, which Table 2 performs the same service for the non-click consonants. A double asterisk notation (**) in the tables indicates a reconstruction that would fit an empty slot in the systemic array, but is not attested as yet by any actual cases, or else has possible reflexes from only one of the three branches. A blank space means no instances of the item have yet been identified in the relevant language. In those instances in Table 2 where the Nama non-click consonant differs from its reconstructed proto-Khoe etymon (in Vossen 1997), the proto-Khoe etymon is placed in parentheses following the Nama reflex.

Table 1. Regular Correspondence of the Click Consonants and Consonant Clusters in SAK

PSAK	!Xoo$	Ju	Nama
*/	/	/	/kh
*#	#	#	#kh
*!	!	!	!kh
*//	//	//	//kh
*Θ	Θ	!	/kh
*/q	/q	/	/kh
*#q	#q	#	#kh
*!q	!q	!	!kh
*//q	//q	//	//kh
*Θq	Θq	!	
*/g	/g	g/	/g
*#g	#g	g#	#g
*!g	!g	g!	!g
*//g	//g	g//	//g
*Θg	Θg	g!	/g
*/G	/G	g/	/g
*#G	#G	g#	#g
*!G	!G	g!	!g
*//G	//G	g//	//g
*ΘG	ΘG	g!	/g
*/h	/h	/h	/h
*#h	#h	#h	#h
*!h	!h	!h	!h
*//h	//h	//h	//h
*Θh	Θh	!h	/h

PSAK	!Xoo$	Ju	Nama
*/q^h	/qh	/'h	/h
*#q^h	#qh	#'h	#h
*!q^h	!qh	!'h	!h
*//q^h	//qh	//'h	//h
* Θq^h	Θqh	!'h	/h
*g/h	/g	g/h	/h
*g#h	#g	g#h	#h
*g!h	!g	g!h	!h
*g//h	//g		//h
**gΘh			
*g/q^h	g/qh		/g
*g#q^h	g#qh	g#	#g
*g!q^h	g!qh	g!	!g
*g//q^h	g//qh		//g
**gΘq^h	gΘqh		
*g/q	/q	g/	/g
**g#q			
*g!q	!q	g!	!g
*g//q	//q	g//	//g
**gΘq			
*/G^h	G/qh	g/	/g
**#G^h			
*!G^h	G!qh	g!	
*//G^h	G//qh	g//	//g
**ΘG^h			

PSAK	!Xoo$	Ju	Nama
*/x	/x	/x	/kh
*#x	#x	#x	#kh
*!x	!x	!x	!kh
*//x	//x	//x	//kh
*Θx	Θx		/kh
*g/x	g/x	g/x	/kh
*g#x	g#x	g#x	#kh
*g!x	g!x	g!x	!kh
*g//x	g//x	g//x	//kh
*gΘx	gΘx	g!x	/kh
*/x’	/kx’	/x’	/
*#x’	#kx’	#x’	#
*!x’	!kx’	!x’	!
*//x’	//kx’	//x’	//
*Θx’	Θkx’	!x’	
*/q’	/q’	/x’	/
**#q’			
*!q’	!q’	!x’	!
*//q’	//q’	//x’	//
**Θq’			
*g/x’	g/kx’	g/x’	/g
*g#x’	g#kx’	g#x’	#g
*g!x’	g!kx’	g!x’	!g
*g//x’	g//kx’	g//x’	//g
*gΘx’	gΘkx’	g!x’	

PSAK	!Xoo$	Ju	Nama
*g/q’	/q’	g	/g
*g#q’	#q’	g#x’	#g
**g!q’			
*g//q’	//q’	(g//)	//g
**gΘq’			
*/’	/’	/’	/
*#’	#’	#’	#
*!’	!’	!’	!
*//’	//’	//’	//
*Θ’	Θ’	!’	/
*/n̥	/n̥		/n
*#n̥	#n̥	n#	#n
*!n̥	!n̥	n!	!n
*//n̥	//n̥	n//	
*Θn̥	Θn̥		/n
*/n	/n	n/	/n
*#n	#n	n#	#n
*!n	!n	n!	!n
*//n	//n	n//	//n
*Θn	Θn	n!	/n
*’/n	’/n	n/	/n
*’#n	’#n	n#	#n
*’!n	’!n	n!	!n
*’//n	’//n	n//	//n
*’Θn	’Θn	n!	/n

PSAK	!Xoo$	Ju	Nama	PSAK	!Xoo$	Ju	Nama
*n/h	/n	n/h	/n				
*n#h	#n	n#h	#n				
*n!h	!n	n!h	!n				
*n//h	//n	n//h	//n				
*nΘh		nΘh	/n				

Table 2: Regular Correspondences of the Non-click Consonants and Consonant Clusters in SAK

(2.A. Non-sibilant obstruents)

PSAK	!Xoo$	Ju	Nama
*b	Ø (b /V_)	b	b (*b)
*b^h		bh	(*p)
*p	Ø		p (!Ora /b/) (*p)
*p^h	Ø	ph	p
*p’	’		p (*p)
**px’	p’kx’		
*d	d (l /V_)	d	d~t (*d)
*d^h	dth	dh	d~t (*d)
*dx	dx	dx	d~t (*d)
*dx’	dt’kx’	dx	d~t (*d)
*t	t	t	d~t (*t)
*t^h	th	th	ts (*th)
*t’	t’	t	d~t (*t)
*tx	tx	tx	d~t (*t)
*tx’	t’kx’	tx’	d~t (*t)

PSA	!Xoo$	Ju	Nama
*g	g	g	g~k (*g)
*G	G	k	g~k
*g^h	gkh	gh	kh
*G^h	Gqh	g	(*g)
*k	k	k	g~k (*k)
*q	q	k	g~k (*q)
*k^h	kh	kh	kh
*q^h	qh	kh	kh
*k’	k’	Ø (*’)	
*q’	q’	k	Ø (*’)
*x’	kx’	kx [x’]	Ø (*’)
*gx’	gx’	Ø (*’)	
*x	x	x	x
*ʔ	’	Ø	Ø (*’)
*h	h	h	h

(2.B. Alveolar and palatal spirants)

PSAK	!Xoo$	Ju	Nama	PSA	!Xoo$	Ju	Nama
*s	s	s	s	*š	s	š	s
*s^{h}	ts	s	ts	*š^{h}	tsh	š	ts
*’s	s	ts	ts	*’š	s	tc	s
*z	dz	z	s	*ž	dz	ž	s
*ts	s	ts	s	*c	ts	tc	ts
*dz	dz	ds	s	*j	dz	dj	ts
*ts^{h}	tsh	tsh	ts	*c^{h}	s	tch	ts
*dz^{h}	dtsh	dsh	s	*j^{h}	dtsh	dch	ts
*ts’	ts’	ts’	s	*c’	ts’	c’	ts
*tsx	tshx	tsx	ts	*cx	tshx	tcx	ts
*dzx	dtshx	dzx	s	*jx	dtshx	djx	ts
*tsx’	ts’kx’	tsx	ts (*tsx’)	*cx’	tshx’	tcx	ts (*tsx’)
*dzx’	dts’kx’	dzx	s	*jx’	dts’kx’	djx	ts

Ordering sound change histories

Despite the relative surface complexity of the correspondences, systemic patternings appear in the sound changes that one can propose to account for the differences. These allow relatively elegant and parsimonious formulations of the underlying sound shifts. The naturalness of the changes that characterize the formulations is in itself a strong indirect suggestion that the correspondence patterns laid out in the etymological dictionary are for the most part valid.

!Xoo consonant sound changes

!Xoõ has been apparently the most conservative in relation to consonant sound change of the three primary languages cited in the dictionary. For the click consonants and clusters, a sequence of two rules can be proposed to account for the attested outcomes. The symbol Q will be used to indicate the click component of a consonant:

1. PSAK [+voice] > !Xoõ [-voice] /#Q_q(’);
2. PSAK [+apirate] > !Xoõ [-aspirate] /QC_, where C = [+voice].

Many of the non-click consonants of PSAK seem to have been preserved in !Xoõ, but two portions of that inventory that have undergone major changes. For the sibilant fricative and affricate consonants and clusters (Table 2.B), four shifts of limited scope, followed by a systemic de-palatalization, can be proposed to account for the attested outcomes of these elements in modern-day !Xoõ:

3. PSAK *c^h > s;
4. PSAK *s^h > c;
5. PSAK *’S > *S (S = *s and *š);
6. PSAK *z > *dz;
7. [+palatal] > [+alveolar].

Most interestingly, !Xoõ dropped all labial obstruent consonants in word-initial position (only [b] persists, and only in intervocalic position). A sequence of two sound changes is sufficient to explain this result:

8. PSAK [+labial/+aspirate] > [+labial];
9. PSAK [+labial/+obstruent] > Ø /#_.

As a consequence, PSAK *b, *b^h, *p, and *p^h > Ø, but *p’ > ’ ([?]), preserving the feature [+glottal] even as it lost its labial locus of articulation.

Ju!hoan consonant sound changes

The sound changes one must propose to account for the Ju/hoan click consonants and consonant clusters are only a few more than for !Xoõ, but the effects of these rules on the systemic distribution of elements was considerably more far-reaching. The major effect was the merging of the uvular feature, both in the click and non-click consonants, with the velar. A de-aspirating shift of very restricted effect, along with a stop de-glottalization shift, can be proposed to have been followed by a general uvular deletion shift. One other very restricted shift of uncertain ordering with respect to the rest would have created the system of click consonants and clusters of present day Ju:

1. PSAK [+ingressive/+voice/+aspirate] > [+ingressive/+voice/-aspirate]
   (PSAK *QG^{h} > PZh *QG, *gQq^{h} > PZh *gQq);
2. PSAK [-contin/+glottal] > [-contin/-glottal] (*t’ > t, *q’ > q, presumably *k’ > k, and *’nQ > *nQ);
3. PSAK [+uvular] > [+velar] (encompassing both clicks and non-clicks];
4. PSAK *n̥Q > *nQ.

In contrast to the major shifts that took place in the spirants in !Xoõ and in the Khoe branch of SAK, Ju/hoan underwent just two small sound changes eliminating PSAK *s^{h} and *š^{h} and also *’s and *’š^{h}, but otherwise apparently preserving the inherited PSAK array of sibilant fricatives, affricates, and clusters:

5. PSAK *S^{h} > S (i.e., *s^{h} > s and *š^{h} > c ([š]);
6. PSAK *’S > equivalent affricate (i.e., *’s > ts and *’š > tc ([č]).

Nama and Khoe consonant sound changes

A much more sweeping recasting of the consonant system took place in proto-Khoe and in its daughter language Nama than in either of the other two branches of SAK. The principle direction of change was toward a loss of articulatory complexity and a greatly simplified inventory overall.

A series of at least seven sound change rules are needed to reduce the PSAK inventory to the proto-Khoe array of click and non-spirant consonants.

1. PSAK *Qg^{h} > PKhoe *Qh,
   *g^{h} > PKhoe *kh; *b^{h} > pre-PKhoe *p^{h};
2. PSAK [+uvular] > PKhoe [+velar] in click consonants;

3. PSAK [-egressive/+glottal/-velar] > PKhoe [-egressive/-glottal/-velar] (this shift would remove the ejectives other than *Q', Qx', *k', *x', *tsx', *dzx', *cx', and *jx' from pre-proto-Khoe);
4. PSAK *(Q) [+obstruent/-glottal] C > PKhoe *(Q) [+obstruent/-glottal];
5. PSAK *Q ([*Qk] > PKhoe *Qx (Nama orthography: Qkh);
6. PSAK *QN > PKhoe *Qn.

Probably also, although the evidence is inadequate for us to be sure,

7. pre-PKhoe *p^{h} > *p.

Most striking of all, the 26 sibilant fricatives and affricates of PSAK were reduced to perhaps five or six in proto-Khoe, *s, *ts, *ts^{h}, *tsx, and *tsx', and possibly *dz, and to two in Nama, *s and *ts. The broad pattern is for the reflexes of most PSAK palatals to have Nama *ts, and most alveolars, Nama *s. But a number of exceptions need to be accounted for. A variety of paths of change could lead to this result. One possible sequences of change would have the following stages:

8. PSAK *'s > pre-PKhoe *č' (this shift would have to have preceded #3 above);
9. PSAK *(t)s^{h}, *š^{h} > pre-PKhoe *č^{h}, *tsx' > pre-PKhoe *cx';
10. pre-PKhoe *ts > PKhoe *s;
11. pre-PKhoe [+palatal] > [+alveolar] (i.e., *č > PKhoe *ts, *j > *dz, *š > PKhoe *s, etc.).

In the proto-Khoekhoe daughter language of PKhoe,

12. pre-PKhoe *dz/*z > proto-Khoekhoe *ts/*s (devoicing shift);
13. PKhoe *ts^{h} > *ts (de-aspiration of affricates);
14. remaining [+uvular] > [+velar];
15. *k' > *x' (unless this shift dates to PKhoe).

Also proto-Khoekhoe created a click from a non-click consonant:

16. *tsx' > */x'.

Nama, a descendant of proto-Khoekhoe, had notable consonant shifts of its own:

17. *(Q)x' > (Q)' ([(Q)?]);
18. *t^{h} > ts;
19. neutralization of the remaining voiced/voiceless distinctions in non-click obstruents.

Tone in PSAK

Tentative and preliminary suggestions about tonal correspondences rest at this point on the fragile basis of evidence comparing two SAK languages, !Xoõ from the Taa-!Ui branch and Ju/hoan from the Zhu branch. In single-mora words, five recurring correspondence pairings have been found, yielding a possible reconstruction of a five-tone PSAK system:

	!Xoõ	Ju	
17. * ˊ	ˊ	ˊ	(and in one case, ˝)
18. * —	—	ˋ	
19. * ˆ	ˆ	ˊ	(and in one case, ˝)
20. * ˋ	ˋ	ˋ	
21. * ˋˋ	ˋ	ˋˋ	

In words of two moras, four tones account for the better part of the evidence, providing we attribute a separate tone to each mora. In !Xoõ one tone—high, low, mid, or falling— continues through both moras; in Ju/hoan irregular matchings of different tones of the two moras appear. The putative reconstructed tones in the left column are no more than highly speculative notions for further investigation:

	!Xoõ	Ju
1. *ˊˋ	ˊ	ˋˋ
2. *ˆ ˋ	ˆ	ˋˋ
3. *ˋ ˋ	ˋ	ˋˋ
4. *-ˋ	-	ˋ ˋˋ ˋˋ
5. *ˊ -	ˊ	ˋ ˊ
6. *ˆ ˊ		ˋ ˊ
7. *ˋ ˊ	ˋ	ˋ ˊ

8.	*-ˊ	-	ˋ ˊ
9.	*ˊˊ	ˊ	ˊˊ
10.	*^ -	^	ˊˊ
11.	*ˋ -	ˋ	ˊˊ
12.	*- -	-	ˊˊ
13.	*ˊ ^	ˊ	ˋˋˋˋ
14.	*^ ^	^	ˋˋˋ
15.	*ˋ ^	ˋ	ˋˋˋ
16.	*- ^	-	ˋˋˋ

a. Ju super-high tone occurs only in a limited set of stem shapes: in click words of the shapes *Cu, *Cau, and *Cai (with one exception, g#a̋); and in non-click words of the shape Ci, where C is only *ts, *ts', or *c (as far as is yet known). It seems therefore most likely a limited, environmentally conditioned development in the Ju group, changing certain reflexes of the reconstructed high tone to super-high; it took place in Ju after the falling tone had merged with the high tone, as the one example of the falling tone going to super-high in Ju reveals.

b. Three unexplained instances of !Xoõ high tone corresponding to Ju low tone in single-mora stems have been noted; otherwise this particular correspondence occurs in cases where morphological operations, usually in !Xoõ, are evident or suspected.

c. Several other infrequent correspondence patterns have been noted, but their attestations are all in cases where morphological operations can be suspected to have caused tone shift.

Being Rigorous in Comparative Khoisan Reconstruction

The consonant reconstructions, in particular, raise two central issues with respect to previous, and very possibly with respect to ongoing, comparative historical work on the South Khoisan (Southern African Khoisan) languages. These issues relate specifically to the establishing of regularity of sound correspondence between click consonants.

One issue relates to the recurring tendency in works on Khoisan relationships to fail to demand full regularity of click correspondences. Somehow, it is thought, clicks might not behave like other consonants and so may have gone through sporadic and individual changes here and there in people's speech. But there is no reason for us to think this way, except for the fact that we are still in the exploratory stage of discovering what the regular correspondences really are. We see partial resemblances between a pair of forms of identical meaning in different languages, and we are tempted to identify them as cognates even though in other cases a different correspondence seems called for.

In number of cases, scholars can find seeming alternate forms of a word in a single language, differing only in their click consonant and thus seeming to give credence to the idea of sporadic click sound change. In each case where I have been able to trace the sources of such forms, it turns out that two separate PSAK roots can be argued to lie behind them. Nama has, for example, two words, !gabo and #gabo, glossed as "to droop." The first can be traced to a root *!Gabo "to be on the verge of falling"; the second to a root *#gabo "to bend, curl." Each meaning depicts an aspect of drooping. They can be understood to be cases of mutual semantic influence between words partially resemblant in pronunciation. They constitute a case like that of ***parameter*** and ***perimeter*** in English. ***Parameter*** has been semantically influenced by the similar-sounding *perimeter* and so is coming in general parlance to mean the "limits or extent of something" rather than identifying a "variable of measurement" relating to that something. In comparative historical reconstruction elsewhere in the world we encounter many examples just like this. They do not constitute grounds in Khoisan, any more than they do elsewhere, for thinking that the requirement of fully regular correspondences can be disregarded in this case. The default assumption when we encounter instances such as these should always be that we just have not yet identified the separate sources of the seeming variants.

A second issue has to do with how we view clicks phonologically. There often seems to be an underlying presumption that clicks form a grouping of sounds like that of labial consonants or velar consonants. Thus sound shifts from one click to another should be relatively common. But in fact clicks have distinctive points of articulation. The distinction between clicks is like that between two stops of different articulatory placement in the mouth. If we encountered, hypothetically, a word in one branch of SAK pronounced *gai and a second word *dai in a different branch of SAK, both with the same meaning, our first inclination would not be to assume that they were cognates. There would have to be some special conditioning factor present before this idea would attract our attention.

Having the same click point of articulation in cognate words in two distantly related Khoisan languages is thus not a trivial correspondence. We expect most often in non-click consonants to find correspondences in position; for example, a labial consonant in a proto-language would more often than not be expected to give labial reflexes in daughter languages. Why should we expect the usual outcomes for click positions to be any different? Just as an original *b might give /v/ in one language and /p/ in another, so a PSAK click consonant such as *!q would be expected to give different consonants, but consonants most likely still pronounced in the same part of the mouth, such as !x (spelled /kh), as it becomes in Nama, and plain *! ([!k]), as it becomes in Ju/hoan.

Consider, in contrast, the case of PKhoe *!x'ao "neck" and !Xoõ #kx'áo "neck." It is highly tempting to jump to the conclusion that these are cognates because the vowels and the click release seem so similar. But we have here clicks of different points of articulation in the mouth, one alveolar and the other palatal. We can find numerous cases supporting the regular correspondence of PKhoe *!x' to !Xoõ !kx', and numerous cases of PKhoe *#x' matching up with !Xoõ #kx', but only this case so far of seeming !x' and #kx'. Are there any special conditioning factors here that might explain such a shift in the point of articulation? Each language has both of these consonants, so there is no gap in the consonant system to be filled by the shift. The vowel environment is not a palatalizing one. So our default assumption here must be that we have a chance resemblance in this instance, and that these words more likely than not had distinct origins.

In addition, a lot of examples are now available of what does happen when clicks change articulation. In general, they do not shift to a different point of click articulation, but instead cease to be clicks. The palatal clicks tend to produce non-click affricates, as Kwadi and East Khoe examples show, and as do Khoisan loanwords into Southern Cushitic in East Africa (Vossen 1997; Ehret, forthcoming). The alveolar clicks most commonly, it appears, drop out, leaving just their velar effluxes behind (e.g., per examples in Vossen 1997 and also in the Taa-!Ui language, //Xegwi, where, for instance, the root *!ui "person" > kwi).

There is one kind of history, though, where we can legitimately expect a click to change its point of articulation, and that is where the inherited system of points of click articulation has been contracted through consonant merger. The original PSAK system had at least five points of articulation—labial, dental, alveolar, alveolar lateral, and palatal—and it may have had a sixth point, alveolar retroflex. (The first five are taken into account in my etymological dictionary.) The notable case in point is the labial position, which was lost in the Zhu and Khoe branches. Different mergers seem to have taken place in the two groups, with the labial going to alveolar (*Θ > *!) in Zhu and to dental (*Θ > /) in proto-Khoe.

If there were once also an alveolar retroflex click in early SAK, as seems very possible (see Starostin in this issue), the most likely merger for it would have been one that maintained the position of articulation—for the simple reason that there were two other alveolar clicks in Zhu and proto-Khoe, *! and *//, for it to fall together with. The case of Ju/hoan g!xòà and Nama //goa-s, both meaning "knee," may attest to just such a process, with the retroflex click going to *! in Zhu and to *// in Nama. According to the rules laid out previously, underlying *gQx yields Nama Qg, but is maintained as *g!x in

Zhu. The vowel sequence *oa is maintained unchanged in both groups (the vowel rules are not given above). So the correspondences are regular in everything but the point of click articulation, and we now have a potential new regular sound correspondence to check out against further examples.

What these cases do not give us is a license to weaken the demands of the comparative method. They are examples indicative, after all, of regular sound change. Undoubtedly irregular sound changes do occur in language history. But they are a small minority within the overarching regular patterns of shift, and even then they are normally fully explainable on the basis of word-specific conditioning factors or linguistic interference. Whether we are dealing with click or with non-click consonants, the requirement is the same—that we demand fully regular sound correspondence from our evidence. There is no basis for treating Khoisan as different in this respect from all the rest of the world's families.

References

Dickens, Patrick. 1994. *English-Ju/'hoan Ju/'hoan-English Dictionary.* Köln: Rüdiger Köppe Verlag.

Ehret, C. 1997. "Transformations in Southern African History: Proposals for a Sweeping Overview of Change and Development, 6000 BC to the present," *Ufahamu* 26, 2: 54-80.0

---------- 2000. "Testing the Expectations of Glottochronology against the Correlations of Language and Archaeology in Africa," in C. Renfrew, A. McMahon, and L. Trask (ed.), *Time Depth in Historical Linguistics*, Vol. 2, pp. 373-399. Cambridge: McDonald Institute for Archaeological Research.

---------- Forthcoming. "Extinct Khoisan Languages in East Africa." In Rainer Vossen (ed.), *The Khoisan Languages of Africa.* London: Curzon.

Starostin, George. 2003. "A Lexicostatistical Approach towards Reconstructing Proto-Khoisan." [In this volume.]

Traill, Anthony. 1994. *A !Xoõ Dictionary.* Köln: Rüdiger Köppe Verlag.

Vossen, Rainer. 1997. *Die Khoe-Sprachen: ein Beitrag zur Erforschung der Sprachgeschichte Afrikas.* Köln: Rüdiger Köppe Verlag.

Appendix: Proposed proto-South Khoisan (PSAK) consonants and consonant clusters

(click consonants and clusters)

*Θg	*/g	*!g	*//g	*#g
**Θg^{h}	*/g^{h}	*!g^{h}	*//g^{h}	*#g^{h}
*Θ	*/	*!	*//	*#
*Θh	*/h	*!h	*//h	*#h
*Θ’	*/’	*!’	*//’	*#’

*ΘG	*/G	*!G	*//G	*#G
**ΘG^{h}	*/G^{h}	*!G^{h}	*//G^{h}	**#G^{h}
*Θq	*/q	*!q	*//q	*#q
*Θq^{h}	*/q^{h}	*!q^{h}	*//q^{h}	*#q^{h}
**Θq’	*/q’	*!q’	*//q’	**#q’

*Θn	*/n	*!n	*//n	*#n
*Θn̥	*/n̥	*!n̥	*//n̥	*#n̥
*Θn^{h}	*/n^{h}	*!n^{h}	*//n^{h}	*#n^{h}
*’Θn	*’/n	*’!n	*’//n	*’#n

*gΘx	*g/x	*g!x	*g//x	*g#x
*Θx	*/x	*!x	*//x	*#x
*gΘx’	*g/x’	*g!x’	*g//x’	*g#x’
*Θx’	*/x’	*!x’	*//x’	*#x’

(non-click consonants)

*b	*d	*dz	*j	*g	*G	
*b^h	*d^h	*dz^h	*j^h	*g^h	*G^h	
*p	*t	*ts	*c	*k	*q	*?
*p^h	*t^h	*ts^h	*c^h	*k^h	*q^h	
	*t’	*ts’	*c’	*k’	*q’	
	*z					
	*s		*š	*x		*h
	*s^h		*š^h			
	*’s		*’š	*x’		
*m	*n					
*m^h	**n^h					
*’m	*’n					

(non-click consonant clusters)

	*dx	*dzx	*jx
	*tx	*tsx	*cx
	*dx’	*dzx’	*jx’
**px’	*tx’	*tsx’	*cx’

(additional possible click consonant clusters in need of further testing)

**gΘq^h	**gΘq’	**gΘq
*g/q^h	*g/q’	*g/q
*g#q^h	*g#q’	**g#q
*g!q^h	**g!q’	*g!q
*g//q^h	*g//q’	*g//q

A Lexicostatistical Approach towards Reconstructing Proto-Khoisan

George Starostin
Russian State University for the Humanities

I. **Introduction.**

I.1. *The current state of "Proto-Khoisan".* Despite all the obvious progress that comparative Khoisan linguistics has undergone over the last half century, no Khoisanologist would deny that crucial questions in this field still remain unanswered. (Basic answers to these questions actually serve as the starting point in any particular area of comparative linguistics.) Not only are we still deprived of a strict and fully credible set of phonological correspondences among the present-day Khoisan languages, we do not even seem sure about whether a genetically related "Khoisan" family actually exists, and whether the "Khoisan" family is any more than a fantasy of some people deluded by the peculiar phonological closeness of most of these languages.

The extreme point of view on this problem, propagated chiefly by the late Ernst Westphal in works such as [Westphal 1965] and [Westphal 1980], is not very popular today, for obvious reasons. While these and other works rightly emphasize the current lack of substantial evidence proving the existence of a genetic relationship between the several established groups of 'Bushman', this by no means gives Westphal a right to claim that such a relationship definitely does not exist. Furthermore, such an approach can hardly be called constructive when it comes to actually explaining what evidence there *is*. This is well understood by modern day scholars; therefore an approach of "moderate scepticism" rather than "decisive denial" is much more popular in Khoisanology today. The difference between the two approaches is summarized well in [Traill 1986], an article in which Anthony Traill both presents the reader with a good selection of comparative material and explains the problems related to its interpretation.

Indeed, if we assume that Khoisan languages are not related, we will be left with a great number of "similarities" between the lexicons of North, South, and Central Khoisan groups (and, to a lesser extent, of Hadza and Sandawe), all of which - including similarities in the basic lexicon and apparent similarities in the morphological inventory - will have to be explained as borrowings or chance resemblances. The number of these similarities may be overrated by some, and there is always a possibility that some of them are indeed the result of lexical diffusion through cultural exchange, but how would that explain, for instance, the use of the same unique form for the 2nd person singular pronoun (PNK **a*, PSK **a*, PCK **c-a* (masc.), **s-a* (fem.; **c-* and **s-* are gender markers)) in all the three major "subgroups" of Khoisan?

On the other hand, assuming there *is* a genetic relationship between these languages, we are faced with the responsibility of establishing the degree of this relationship for every individual subgroup, as well as (more importantly) a system of strict phonological correspondences between the languages supposed to be related. Here there is no consensus among researchers and, in fact, not much work has been done so far in either of these directions. The existence of the so-called "North Khoisan" (or "Žu"), "South Khoisan (or "Taa-!Wi"), and "Central Khoisan" (or "Khoe") language families is hardly debatable, considering the numerous isoglosses and grammatical similarities within each of the three; but the 'arboreal' connections between these three families are not easily defined, not to mention their ties with Khoisan "isolates" - Hadza, Sandawe, #Hoan, and Kwadi. Even assuming that all these branches go back to a common Proto-Khoisan ancestor, are they all equally distant from PK or can they be first reduced to several larger families?

As for the problem of sound correspondences, what little work has so far been done on this problem should mostly be credited to Henry Honken. In several articles (see especially [Honken 1988] and [Honken 1998]) he has made quite a few interesting observations on the possible types of sound correspondences in Khoisan. His classification of these correspondences into "sporadic," "quirky," "conservative," and "classical" ([Honken 1998]), which might look like a joke in theory, should actually be taken very seriously. In fact, the main point of this classification - namely, that phonological correspondences between Khoisan languages are much more complicated than the overall similarity of the existing phonological systems suggests - turns out to be crucial in understanding the very essence of Khoisan historical phonetics. On the down side, Honken offers no stable system of correspondences for Khoisan, and the principles he uses for comparison are altogether unclear. As useful as his works are, they present no more than isolated snippets of what could possibly be called "reconstruction of Proto-Khoisan," and differing in quality at that. Thus, the compared material often drastically alternates between "basic" lexicon and "cultural" lexicon, meaning that quite a few comparisons could easily be discarded as potential borrowings, especially if they are not confirmed by similar correspondences in the basic vocabulary.

Christopher Ehret, who is currently working on a comparative dictionary of Khoisan, suggests a somewhat different approach. (An early sketch of his ideas on Proto-Khoisan can be found in [Ehret 1986], and a much more exact and detailed description of Proto-Khoisan phonetics is present in the current volume [Ehret 2003]). Unlike Honken, Ehret has a work-in-progress system of correspondences between North, South, and Central Khoisan, which is largely based on the "one-to-one" principle, i.e. a system where phonemes correspond to each other in a strict and simple way (dental clicks to dental clicks, alveolar clicks to alveolar clicks, etc.). Only where a feature or a set of features is distinctive of only one branch are "non-trivial" correspondences allowed, and even then the historical processes are more or less predictable (i.e., the development of uvular effluxes into regular velar effluxes in North and Central Khoisan - where uvular effluxes are not found - as opposed to their preservation in South Khoisan). As rigorous and formal as the ensuing results are, this inevitably means neglecting a lot of work conducted by Honken, for example, automatically discarding all of his "sporadic" and "quirky" correspondences as mere chance resemblances. Another problem is that the tighter the phonological limitations on possible correspondences, the looser we find the semantic criteria of data selection. Among Ehret's comparisons only a minor handful actually comprises "one-to-one" semantic matches among the basic lexicon, which inevitably raises the question of possible oversights in the compared material.

All in all, while both Honken and Ehret's contributions to comparative Khoisanology must necessarily be taken into account by anybody interested in reconstructing Proto-Khoisan, it cannot at the present time be said that either of these gives us a clear perspective on the nature of the former. One of the main reasons for that appears to be the lack of a proper "starting point" for delving into Proto-Khoisan. Before dealing with a language family as such (i.e., a group of languages historically descended from one common ancestral language), one needs to *prove* that such a language family actually exists. Traditionally, a language family is said to be proved if we have a regular set of phonological correspondences working on a large percentage of the basic vocabulary of all the languages in the said family, preferably supported by similarities in the languages' morphological systems as well. Nothing of the kind has been proposed so far either by Honken or Ehret; of the two, Ehret comes closer to fulfilling these demands, but given the extreme semantic fluctuation of a large part of his comparisons, one can have serious doubts about the validity of many of the suggested correspondences, or, at least, one may reasonably assume that there may be a large number of additional correspondences that have not been spotted by Ehret.

I. 2. *Lexicostatistical principles of analysing Khoisan data.* This article will try, in a way, to combine both the formal approach chosen by Ehret and the "observationist" approach of Honken by using both in a lexicostatistical analysis of available Khoisan data. It is normally assumed that lexicostatistics should be conducted after the establishment of the system of correspondences, not before, but occasionally it helps to reverse the procedure, since lexicostatistics can be an important tool in deciding upon the possible correspondences - that same "starting point" that we are looking for. Using the strict lexicostatistical approach, we can single out cases of possible correspondences with less risk of confusing them with results of borrowing, and then verify their status by trying to find data outside the Swadesh 100-wordlist that confirms them. It should be noted that although such an approach does not exclude the possibility of taking chance resemblances for regular correspondences, limiting ourselves to the 100-wordlist severely limits this possibility. And, of course, the lexicostatistical test will at the same time help us build a more detailed classification of Khoisan.

So far, a serious lexicostatistical analysis of Khoisan has only been conducted by Bonnie Sands in [Sands 1998], where it serves as one of the several methods of evaluating Khoisan lexical data on the subject of possible genetic relationship. Apart from the problem of Sands choosing an alternate 100-wordlist, specially "adjusted" for Khoisan realities (with words such as 'elephant' and 'giraffe' making the list among others), the main flaw of her approach is an exaggerated demand for similarity; thus, only words where click influxes in different languages coincide are counted among possible correspondences. Yet no one has ever proved that clicks cannot change from one branch of Khoisan to another; in particular, such an approach would be in direct opposition to whatever results Honken has achieved in his latest works. In fact, if it were to be found out that clicks *never* correspond directly to one another (i.e. North Khoisan dental clicks never correspond to South or Central Khoisan dental clicks, etc.), such an approach would be downright wrong.

The following limitations will be imposed on the lexicostatistical analysis data below to make it more reliable:

1) For smaller language groups whose existence is usually not contested (namely, North Khoisan, the two subgroups of South Khoisan and the two subgroups of Central Khoisan) I will try to give the proto-form instead of concentrating on one or two languages. This is practical in that for many individual languages, especially extinct Southern Khoisan ones, the 100-wordlist is far from available, yet collective analysis of available data can still yield a possible intermediate proto-form. Sometimes such an approach can lead to excessive synonymity (more than one possible proto-form for one word), but a limited amount of synonymity can actually be permitted in lexicostatistical/glottochronological calculations. On the other hand, it helps avoid such "uncomfortable" synonymity as, for instance, met in Žu|'hoan designations of body parts, where we often meet two terms for one part - one an original North Khoisan root (*!kxá* "heart", *cʔí* "mouth") and the other most probably a Khoekhoe borrowing (*#áó* "heart", *kxám* "mouth"), confirmed by the fact that while the former roots yield numerous parallels in other North Khoisan dialects, the latter are rather a peculiarity of Žu|'hoan and are, for instance, totally absent among North Khoisan forms in Dorothea Bleek's comparative vocabulary. Thus, even if based exclusively on Žu|'hoan lexical data we cannot state with certainty what the main Žu|'hoan words for "heart" or "mouth" are, we can definitely assert that the Proto-North-Khoisan roots were **!kxa* and **cʔi.*

2) The most important thing, of course, is to determine what exactly constitutes a match in the 100-wordlist and what does not, i.e., a proper method of setting up possible phonetic correspondences between languages. Here the following rules will be proposed:

(a) word A in language X will be considered a match with word A_1 in language Y if their segment structures (apart from possible minor vocalic differences) coincide, unless it can be *proved* that at least one of the direct correspondences between A and A_1 does not really exist, or it can be *proved*, at least in an indirect way, that word A is a borrowing from Y or *vice versa*. *Example*: Proto-North-Khoisan **/am* "sun" will be considered as a match with Proto-Central-Khoisan **/am* id., since the two words match in consonant structure; it has not been proved that the two cannot correspond; and finally, there is no significant evidence that the North Khoisan proto-form was borrowed from Central Khoisan, or *vice versa*;

(b) if the segment structures of A and A_1 do not coincide, they will still be considered a match if there is at least *some* significant evidence that they might possibly correspond to each other. This evidence has to consist of a certain number of additional examples (B/B_1, C/C_1, etc.) containing the same correspondences as found between A and A_1. (This number may range from one to several, depending on the relative frequency of the involved phonemes and the supposed level of proximity between the languages in question.) In order to drastically reduce the possibility of chance similarities, this evidence has to belong to the basic lexicon section and boast strong semantic correlations between the compared forms (preferably cases where the meanings of B and B_1, etc., fully coincide); one or two strong supportive examples of that kind can often be more convincing than a dozen weak examples with dubious semantic shifts. *Example*: Proto-North-Khoisan **#àʔú* "cold" corresponds to Proto-Taa **//âʔu* id., because of the existence of such other examples as PNK **#nà̯m* "frog" - PT **ʔ//nā̯ʰm* id., etc. (for more of these, see II.15);

(c) considering the huge number of clicks in many of the compared languages, it is wiser to treat click effluxes and influxes separately when dealing with possible correspondences. A particularly rare click, like the ones with effluxes *-Gh-* or *-ŋ̥-* in !Xóõ, might not yield any obvious correspondences in other languages when taken all by itself; yet separating it into the efflux and influx part and trying to determine separate correspondences for each will most certainly produce better results, as this effectively provides us with more possible comparative evidence for every type of sound. Thus, word A with the structure IEV (where I = click influx, E = click efflux, V = the rest of the stem) in language X is a match for word A_1 with the structure I_1E_1V (where I_1 and E_1 are a click influx and a click efflux *different* from I and E) in language Y, if it can be shown, on the basis of supporting evidence, that I is a possible match for I_1 and E is a possible match for E_1, even if there is no other convincing example of the correspondence IE/I_1E_1. *Example*: Proto-North-Khoisan **!̥ùʔúru* "claw, nail" corresponds to Proto-Taa **//qûrV* id., even if there are no other examples of the correspondence "PNK **!̥* - PT **//q*". There are, however, examples of the correspondence "PNK **!̥* - PT **//*" and the correspondence "PNK zero efflux - PT **-q-*" (see II.13);

(d) "tentative" matches can be established if the suggested correspondence is not confirmed by additional data, yet is not contradicted by any other correspondence. Thus, click efflux E in language X can be said to correspond to click efflux E_1 in language Y even if there is only one word with click efflux E in language X corresponding to one word in language Y, provided that no other efflux of language Y, besides E_1, can correspond to efflux E. Note, however, that in order to have a relatively high probability of relationship, efflux/phoneme E must not be too phonetically divergent from efflux/phoneme E_1. Considering the hugeness of the phonological inventory of Khoisan languages, this is a very important addition, which can certainly mean having several erroneous comparisons, but also allows us to track down multiple isoglosses that would be "lost" otherwise. *Example*: Proto-Taa **/qʔan* "heart" can be matched with

Proto-North-Khoisan **!kxá* id., because no other correspondences in North Khoisan to PT **-qʔ-* have been established (at least, no other correspondences with strong semantic resemblance), and the two effluxes are quite close phonetically (both are ejective, and in addition, North Khoisan possesses no uvular effluxes whatsoever);

(e) if neither of the four rules above can be shown to apply to the word pair A vs. A_1, they do not constitute a match, no matter how close their phonetic likelihood may be. Example: PCK **//ʔo* "to die" cannot be matched with Sandawe *ƛa-si* id. despite the presence of a lateral consonant in both words, because no other convincing examples with the same correspondence have been found. This is a very important addendum, since, for instance, the presence of the same click influx in two roots of different subgroups often gives a potentially misleading impression of genetic relationship - even if the efflux and the rest of the stem cannot be proven to match at all.

This means that the important thing for us in this procedure is not so much *likelihood* as it is *correspondence* confirmed by supporting material. (As I am going to show, quite often the compared forms are not going to look similar at all.) Note, however, that "correspondence" does not necessarily presuppose "one-to-one correspondence"; such a presumption would obviously be biased, and might prevent us from recognizing many important matches within the 100-wordlist. This work, then, is being done based on a presumption of *probable non-simplicity* - i.e. that we are by no means forced to limit ourselves to comparing material of the rule (a) variety when looking for potential correspondences. It should be emphasized once again here that the main goal of this work is *not* to establish a self-sufficient system of regular correspondences, which is at the present time a near-impossible task; the main goal is to use a more or less formalized method to track down and fix *as many types of potential correspondences as possible*, which will be systematized and categorized later on the basis of additional data.

I. 3. *Data sources.* For space reasons, I will refrain from indicating the exact data source for every Khoisan form quoted below; the absolute majority of the data actually comes from a list of predictable sources quoted in this section. To avoid any confusion, I have also chosen, where possible, to unify the transcription, which sometimes means significantly changing the graphic inventory of certain sources; a list of transcription signs used in this article can be found at the end. The only language with "non-unified" transcription is Nama, since it is the only Khoisan language that currently has something at least remotely resembling a commonly accepted orthographic norm. Abbreviations listed below will be used throughout the work, with the preceding P standing for Proto (e.g. NK = North Khoisan, PNK = Proto-North-Khoisan).

North Khoisan (NK). The Žu|'hoan form is quoted according to [Dickens 1994], with unified transcription (Dickens' designation of the Žu|'hoan affricate system can be particularly baffling, even for those familiar with the field). The major source for other dialects, such as //Au//en and !O!Kung, is [Bleek 1956]; of more recent publications, [Snyman 1997] is of exceptional interest, but, unfortunately, it only provides a very limited amount of data.

South Khoisan (SK). This subgroup is traditionally assumed to be subdivided into two relatively distant branches, Taa (T) and !Wi (Kw), and since there are some really interesting (and complex) lexical differentiations between the two even within the 100-wordlist, I have preferred to deal with them separately. For Taa, the main source is Anthony Traill's near-exhaustive dictionary of !Xóõ [Traill 1994], as well as additional Masarwa and |Nu//en data from [Bleek 1956]. For !Wi, a supposedly extinct language branch, the only major source is [Bleek 1956], with a few additional forms from [Bleek 1929].

Central Khoisan (CK). Again, this large group shows a clear split into a "Khoekhoe" (KK) branch

(Nama and !Ora) and a "Non-Khoekhoe" (NKK) branch (everything else), with big lexical differentiations between the two. They are thus treated separately. Central Khoisan is the only division of Khoisan for which we have an "official" reconstruction of the phonological system, published in [Vossen 1997] (the older reconstruction by Kenneth Baucom [Baucom 1974] has to be considered obsolete); thus, where possible, I cite Rainer Vossen's reconstructions for both Proto-Khoekhoe and Proto-Non-Khoekhoe forms. In several cases where Vossen does not provide a reconstruction, I offer one myself based on other data sources, such as [Haacke 1998] for Nama, [Meinhof 1930] for !Ora (Korana), [Barnard 1985] for Naro, [Tanaka 1978] for |Gwi and //Ganakwe, and [Dornan 1917] for Hietšware.

#Hoan. The data on this supposedly isolated Khoisan language still remains largely unpublished; however, unlike the other "isolate", Kwadi, #Hoan has at least been seriously studied by several researchers, and some #Hoan data, enough to fill in about three quarters of the 100-wordlist, can be found in works such as [Bell-Collins 2001], [Collins 2001], [Collins 2001a], [Collins 2001b], [Gruber 1975], [Traill 1973], and others; large chunks of #Hoan lexical data, collected by Chris Collins and others, can also be found at the site maintained by Cornell University at http://instruct1.cit.cornell.edu/courses/ling700/.

Sandawe. The main source is [Kagaya 1993]; for additional safety, most of Kagaya's data have been checked by me against Otto Dempwolff's earlier vocabulary of Sandawe [Dempwolff 1916].

Hadza. No major Hadza vocabulary is yet available, unless one counts Dorothea Bleek's data in [Bleek 1956] (which is usable, but has always to be checked against newer, more phonetically exact sources for safety). For lexicostatistical comparison, of crucial importance is the Hadza 100-wordlist provided by Bonnie Sands in [Sands 1998]; some of the words that are not present in her version I have managed to discover browsing through unpublished field materials collected by Archibald Tucker and Derek Elderkin (courtesy of H. Fleming).

I. 4. *Notes on intermediate reconstructions.* As I have already mentioned above, the present work uses intermediate reconstructions rather than separate languages for lexicostatistical analysis. This does not, of course, apply to "isolates" like #Hoan, Hadza, and Sandawe, and in a couple of cases the "reconstruction" in question is little more than a possibly slightly modified - or not modified at all - form of one major representative of the group, depending on the number of languages and on the degree of their proximity. Speaking in individual terms, this is what must be mentioned specifically:

a) North Khoisan. This is obviously a very young language branch, with a high level of mutual intelligibility between all of its speakers. The dialect data presented in [Snyman 1997] clearly shows that there are few phonetic differences between the actual dialects (at least on the segment level; Khoisan tonology and prosodics are so complex and vary so seriously between even closely related languages that we do not have the time, nor the ability to raise that issue here). The main differences are to be found in the affricate/fricative system, which is very large in NK (apart from Hadza, it is the only branch of Khoisan that differentiates between the hissing and the hushing series), and apparently rather unstable, judging by the extremely non-systematic and numerous correspondences. The exact number of affricates in PNK, and their reflexation in daughter dialects, is yet to be established; Snyman's examples are too few to base any decisive conclusions upon them. For the moment, we are taking the Žu|'hoan system as described by Patrick Dickens as a hypothetic "substitute" for the Proto North Khoisan one.

One extremely important thing about North Khoisan is what could be called the "fifth click problem". It has been noted by several researchers, including Snyman himself, that some North Khoisan words beginning with the alveolar click ! tend to preserve it in all the dialects, while certain other words tend to substitute it - either with the lateral click //, or with a special type of click articulation, for which C. M. Doke had much earlier proposed the term "retroflex" [Doke 1925, p. 148]. The "retroflex"/alveolar/lateral trifurcation of the click seems to be more or less regular, depending on the particular dialect (according to Snyman, Northern dialects tend to have the lateral variant, Southern dialects tend to have the alveolar variant, and Central dialects fluctuate between the retroflex and lateral variants), but the reasons for this trifurcation are still obscure; Snyman's attempt [Snyman 1997, p. 35] to explain it through prosodic influence ("retroflexization" before a high tone) can hardly be called satisfactory, both because of a large number of countercases and the unclear character of the mechanism of phonetic change.

The problem, however, can be successfully eliminated - on the Proto North Khoisan level at least - if we suppose that the "retroflex" articulation of the "trifurcated" click is actually original. Reconstruction of five, instead of four, clicks for North Khoisan, at this point seems not only the best, but the _only_ way to deal with the problem based on the classic comparative method. The original alveolar click, then, remains unchanged in all NK dialects, while the original retroflex click is preserved in but a few, having merged with either the alveolar one or the lateral one in most others in a general simplification-of-the-system process. We can even define quite a few minimal pairs for PNK, such as _*!gu_ "belly, stomach" (//Au//en _!gu_, Ju|'hoan _!gú_, !O!Kung _!gu_) vs. _*!̡gu_ "water" (//Au//en _!gu, //gu_, !Kung _!̡gu_, Žu|'hoan _!gú_, !O!Kung _//gu_), etc. Obviously, the downside of this decision is that it "burdens" us with yet another opposition to be explained on the Proto-Khoisan level, but on the other hand, it might just as well provide extra insight into the original phonological system.

The hypothetical PNK form, then, can be described as "The Žu|'hoan form with the alveolar click replaced with a retroflex where necessary _and_ possible". Unfortunately, too often the only form in our possession is the Žu|'hoan one, and since Žu|'hoan always replaces the original retroflex click articulation with the alveolar one, without additional dialectal data it is impossible to determine whether it goes back to _*!_ or _*!̡_.

b) South Khoisan. The genetic unity of this branch is quite obvious and can easily be seen from Westphal's lexical data in [Westphal 1965]; a detailed study of the problem with positive results can be found in [Hastings 2001]. The Proto-Taa form is essentially the !Xóõ form as given in [Traill 1994], possibly with a trimmed derivational suffix to emphasize the exact form of the root; Bleek's data on Masarwa and |Nu//en add little, and what little differences there are can actually be due to errors in transcription. The same problem, but in a much worse form, arises when we attempt to reconstruct Proto-!Wi: there is not one fully reliable (in terms of transcription) source of data for any of the languages in this subgroup, and any hypothetic reconstruction of a Proto-!Wi form, based primarily on |Xam, #Khomani, and Batwa (//Xegwi) material, would have to be "twice hypothetic" because of poor transcription quality. Yet, as will be seen below, this data should not be neglected altogether, as in certain cases it can provide valuable insight into some of the processes in Khoisan historical phonetics. However, out of all the reconstructions below, the Proto-!Wi one is, without a doubt, the most questionable one.

c) Central Khoisan. As mentioned above, most of the intermediate reconstructions of both PKK and PNKK are either taken directly from [Vossen 1997], or based upon the correspondences set up by Rainer Vossen in that work. The few cases of disagreement with or modification of Vossen's correspondences (e.g. the treatment of uvular effluxes, etc.) will be discussed specially.

II. Lexicostatistical data with commentary.

Preliminary notes. Obviously, these wordlists do not pretend to be a hundred percent correct. For some subgroups in some cases it is impossible to reconstruct any proto-form, due either to lack of data or to the existence of too many candidates for one position. In most cases, however, we can come up with a rather clear idea of what the proto-form (or two synonymous proto-forms) actually looked like. For suppletive verbs with different singular and plural stems in PNK, PT, PKw, #Hoan, and Sandawe, the singular action stem is taken as the default one.

II.1. "ALL":

A) PNK ***wV**(//Au. *wasi*, Žu. *wa̭ᵑ-sì*, *wè-šè*, !O. *wešeše*); PKK ***hoa** (Nama *hoa*, !Ora *hoà*); PNKK ***we** (Naro *we-*, |Gwi *we-ha*, //Gana *we-kae*); Hadza **waina**; B) PT ***kU** (!Xóõ *kōo kàʔàᵑ*, Masarwa *kuka*); PKw ***ku** (|Xam *ku*, //Ng *kw-a*, Seroa *ku*); C) Sandawe **chia**.

Notes. The NK, CK, and Hadza roots constitute a match according to rule (a), if we suggest that Hadza *-na* is an affix. It is unclear whether the initial aspiration in Khoekhoe is original (and the PK root should look something like **hwe-*) or prothetic, but it is hardly dubious that it belongs together with the rest of Central and North Khoisan forms. The structure is actually unusual, as there are very few roots in any Khoisan branches (outside of Hadza and Sandawe) starting with a labial w-, and the root for "all" is the only one that shows this feature in several branches at once, confirming the connection. It is theoretically not excluded that SK **kU* < **(h)wV*, but this assumption is unverifiable, so no sure match can be postulated here.

II.2. "ASHES":

A) PNK ***to̭** (//Au. *tɔ*, Žu. *to̭*, !O. *tā̭otā̭o*); PKK ***thao** (Nama *tsao-b*, !Ora *thao-b*); PNKK ***thau** (Naro, //Gana, #Haba, Danisi *thàú*); B) PNK ***#goa** (!O. *#gwa*); PT ***#goa** (!Xóõ *#gòa*, Mas. *//gwa*); PNKK ***#goa** (//Ani *#gôà*, Buga *#nòá*, |Ganda *#nòà*, |Gwi *#gûà*, //Gana *#gôà*, |Xaise, Kua *ʒóá*, Cara *ʒóàà*, Tsixa, Danisi *djûà*, Tsua *ʒùàá*, Hie. *ǯoa*); C) PKw ***!ui** (|Xam, #Kho. *!ui*); D) Sandawe **!ʔupha**; E) Hadza **ho-çọ**.

Notes. Two different Proto-Khoisan roots are obvious here; judging by the fact that the meaning of the first one tends to develop into "flame" (in //Ani and Kxoe in particular), while the second one often means "clay, mud" (Nama *#goa-b*, etc.), it is possible that the original differentiation had been along the lines of "hot ashes" and "cold ashes, dust". The correspondence between PNK **t* and PCK **th* is surprising (considering that both distinguish between aspirated and non-aspirated phonemes), but cf. also cases like PNK **tama(h)* vs. PCK **thama* "tsamma melon" which show this is possible.

Root (b) serves as the regular form for "ashes" in !O!Kung (the "North" cluster of NK dialects); in Žu. the meaning of *#goah* is 'soap', with a possible influence of (or even direct borrowing from) Khoekhoe.

One cannot theoretically exclude that PKw **!ui* and Sandawe *!ʔupha* are related, but apart from the same click influx there is little to support that hypothesis, which leaves unexplained the lack of glottal stop efflux in PKw and, especially, the second syllable in Sandawe.

II.3. "BARK (of tree)":

A) PNK ***//no̭ʔòrV** (Žu. *//no̭ʔòrò*, !O. *//nuli*, *!nuli*); B) PT ***gu** (!Xóõ *gúle*, pl. *gún*, Mas. *gule*, |Nu. *!gum*); (?) PKw ***(!)go** (//Ng *!go*); PNKK ***gu-re** (|Gwi *gure*, //Gana *gure*, Hie. *ǯore*); C) PKK ***soro** (Nama *soro-b*, !Ora *soró-b*); D) PNKK ***//kxúᵑ** (//Ani, Kxoe, |Ganda *//kxúᵑ*, Buga *//kxǔᵑ*, Naro *//kxṳ́ᵑ*, #Haba *//ʔṳ̌ᵑ*, |Xaise, Cara, Tsixa, Kua *//ʔúᵑ*); E) Sandawe **ʎa̭**; F) Hadza **he-k̭wá**.

Notes. The only root here that has a fairly wide distribution is _*gu_ or _*go_, possibly _*gurV_ if the suffix alternation in !Xóõ is a secondary morphological process cutting across the original root structure. However, the root is rather local in CK (only a few languages), and its presence in !Wi languages is equally dubious (although there are certain cases in Bleek's dictionary where clickless roots are being transcribed with "false" initial clicks, the main problem here is that the root is only found in one language).

On the other hand, it is the only root, apart from the isolated PCK _*//kxúᵖ_, that cannot be traced back to a different meaning; PNK _*//nọʔòrV_ begs for comparison with PT _*//Gụ̀lV_ "to peel, strip, remove bark", while PKK _*soro_ goes back to PCK _*córó_ "shell, pod". The Sandawe and Hadza forms look slightly similar, but there is no serious evidence confirming the _λ̣_ - _ḳ_ correspondence, not to mention any possible ties with the other languages' data.

II.4. "BELLY":

A) PNK ___*!gú___ (Žu. _!gú_, !O. _!gu_, //Au. _!gu_ "body"); #Hoan ___!o___; B) PT ___*!hūma___ (!Xóõ _!hūma_); C) (?) PKw ___*!autu___ (|Xam _!autu_); D) PKK ___*!na___ (Nama _!na-b_, !Ora _!nà-b_); E) PNKK ___*/a~*/na___ (//Ani _ᵖnǎ_, Kxoe, Cara, Danisi, Cua _/ǎ_, Buga _/á_, |Ganda, |Xaise, Kua _/à_, Naro, |Gwi, //Gana _ᵖnâ_, #Haba _ᵖnà_, Deti, Tsixa, Tsua _/â_); F) Sandawe ___λ̣abiso___; G) Hadza ___ho-ča___.

Notes. A good match between NK and #Hoan; the correspondence "NK voiced efflux - #Hoan zero efflux" can be found in at least several other good examples, e.g. PNK _*!̣gàʔámá_ "to enter" - #Hoan _!ạam_ "to enter (pl. stem)", etc. On the other hand, tracing a match between NK and the several isolated SK forms would be more risky. Theoretically, |Xam _!autu_ and !Xóõ _!hūma_ might be related, but that would require explaining the different suffixes.

It is also tempting to compare the Khoekhoe and Non-Khoekhoe forms, but no sufficient evidence exists to suppose a click replacement from dental to alveolar (or vice versa) within CK. We will have to assume that the two roots are unrelated.

II.5. "BIG":

A) PNK ___*!̣naʔà___ (//Au. _!na_, !Kung _!!nʔa_ (Doke), Žu. _!naʔàᵖ_, !O. _//na_); B) PT ___*!xa___ (!Xóõ _!xaV_, Mas. _!xa-i_); C) (?) PKw ___*!ui___ (|Xam _!ui-a_); D) #Hoan ___//nam___; E) PKK ___*kai___ (Nama _kai_, !Ora _kai̋_); PNKK ___*kai̋___ (Naro _kai̋_, Deti _kai̋_); F) Sandawe ___baʔe___, Hadza ___pakapaʔa___.

Notes. The NK and #Hoan forms are possibly related, but so far no other convincing examples of the PNK _*!̣_ - #Hoan _*//_ correspondence have been found (running ahead, I should note here that click influx correspondences between NK and #Hoan, apart from those involving the #Hoan labial click, happen to be extremely stable, with the retroflex click regularly corresponding to the #Hoan alveolar click). The CK forms are obviously connected with !Xóõ _kái_ "to grow".

The Hadza and Sandawe forms are obviously related (the Hadza form looks like a composite, with the second root the same as _pa(ʔ)a_ "great, old"), although the voicing (or devoicing) is unclear. Strictly speaking, it could be discredited due to "irregular" correspondences, but the root _*PV_ meaning "big" or "many" is more or less a global etymology, and this gives additional support to this particular matching.

II.6. "BIRD":

A) PNK ___*cʔàmà___ (//Au. _cama_, Žu. _cʔàmà_, !O. _caba_, _cama_); #Hoan ___chạma___; B) PT ___*/gū___ (!Xóõ _/gūhʔu_, |Nu//en _si-/kou_, (?) Mas. _ši-//gu_); PKw ___*/u-i___ (|Xam, //Ng _/wi_, //Kxau _/hwīᵖ_); C) PKK ___*kxani___ (Nama _ani-_, !Ora _kxàní-s_); D) PNKK ___*ʒàrá___ (//Ani, Buga, Naro, #Haba, Danisi _ʒará_, Kxoe _ʒǎdȧ_, |Ganda, |Gwi _ʒárá_,

//Gana *ʒárà*, |Xaise, Tsua *ʒérà*, Deti *ʒàrà*, Tsixa *ʒìrá*, Kua *ʒérá*, Hie. *zera*); E) Sandawe ***thui***; Hadza ***thithi-***.

Notes. Another clear match between NK and #Hoan (it is quite possible that the preglottalization of the affricate in NK correlates to the pharyngealized vowel in #Hoan; aspiration can be neglected because of frequent variation of aspirated/unaspirated variants in North Khoisan dialects). Some Taa dialects display a strange composite form for bird (**si-/gu*), but the main root is obviously the same for most SK languages.

The original PCK root for "bird", **ʒàrá*, was replaced in PKK by **kxani*, originally "vulture" as shown by Non-Khoekhoe data.

Hadza and Sandawe forms closely match in the consonantal structure; the Hadza form is most certainly a reduplicated stem, possibly with the reduction of the original diphthong (**thuithui-* > *thithi-*).

II.7. "BITE":

A) PNK ****!nai*** (//Au. *!ne*, Žu. *!na̋i*, !O. *!nai*); B) PT ****síʔi*** (!Xóõ *síʔi*, (?) |Nu//en *tseja //ai*); PKw ****cʔi*** (|Xam, //Ng, #Kho. *ci~cʔi*); Sandawe ***/ʔi^{ŋ}-khe***; C) #Hoan ***!gai***; D) PKK ****Pa*** (!Ora *bá*); PNKK ****pá*** (//Ani, Kxoe, Buga, |Ganda, |Gwi, //Gana, |Xaise, Deti, Cara, Tsixa, Danisi, Kua, Tsua *pá*, Hie. *pha*); E) Hadza ***kaʔ-ƛ̣é-***.

Notes. As much as the NK and #Hoan forms are similar, it is very hard to explain the nasal efflux in PNK - there is no other clear-cut example where it would appear so thoroughly unmotivated. Instead, I strongly suspect that #Hoan *!gai* is the same root with *!gai* "snake", corresp. to PNK **!̣gai* "puff-adder", a meaning shift all too common in other Khoisan branches.

This shift is, in fact, present in Proto-Taa, where the root **síʔi* means both "to bite" and "snake"; and similarly, Sandawe */ʔi^{ŋ}-khe* "to bite" looks very much like a derivation from */ʔi^{ŋ}* "snake". The resemblance between PSK **cʔi* and Sandawe */ʔi^{ŋ}* is obvious, and the correspondence between the SK affricate and the Sandawe dental click reappears one more time in the list (item II.16), which makes the probability of the forms being related even higher. This is, then, a definite match.

II.8. "BLACK":

A) PNK ****žó*** (//Au. *žɔ*, Žu. *žó*, !O. *ǯo*); B) PT ****#a[-ʔna]*** (!Xóõ *#áʔná*, (?) Mas. */kxa*, |Nu//en *#ana*); #Hoan ***#kxau***; C) PKw ****!Ue*** (|Xam *!ue-n*, //Ng *!oe*); D) PKK ****#nu*** (Nama *#nu*, !Ora *#nù*); PNKK ****#nú*** (//Ani, |Gwi, //Gana *#^{ŋ}nú*, Kxoe, |Ganda, Tsixa *#^{ŋ}ú*, Buga *#^{ŋ}ǔ*, Naro *#^{ŋ}nǔ*, |Xaise *nʒu*, Deti *jǔ*, Cara *jú*, Danisi *ndú*, Kua *ʒú*, Tsua *dú*, Hie. *ǯu-nje*); E) Sandawe ***ḳaŋḳara***; F) Hadza ***tịčị-***.

Notes. The match between PT and #Hoan, with an "extra" velar affricate efflux in #Hoan, is possible; cf. examples like #Hoan *//kxao* "to chop" - PT **//a̰*, pl. action **//a̰o* id; also "small" (II.77). If the mysterious Masarwa form (*/kxa*) belongs here indeed, with an incorrectly transcribed click influx, it may represent an earlier variant of the PT form.

Unfortunately, no other matches are found. The resemblance between forms like !O. *ǯo* and Hie. *ǯu-nje* turns out to be false, as the latter goes back to the regular **#nu* proto-form, with secondary affricativization of the palatal click; no such development can be established for NK. It is interesting to note that both the Sandawe and the Hadza form may be reduplicated (*ḳaŋḳara* < **ḳarḳara*, *tịčị-* < **čịčị-*), but that does not mean their segment structure is actually comparable.

II.9. "BLOOD":

A) PNK ****/ʔŋ́*** (//Au. */ʔ^{ŋ}*, Žu. */ʔŋ́*); #Hoan ***/qʔi***; PKK ****/ʔao*** (Nama */ao-b*, !Ora */ʔau-b*); PNKK ****/ʔáò*** (all languages exc. Kua, Tsua, and Hie. have */ʔáò*); B) PT ****!na̰*** (!Xóõ *!nà̰a*, Mas., |Nu//en *!na̰a*); C) PKw ****//xau***

(|Xam _//xau-ken_, //Ng _//xau_, |Auni _//xauʔu_); D) Sandawe __//ʔeḳa__; E) Hadza __ʔáthaʔma-__.

Notes. The SK forms are unclear (unless by some chance PKw _*//xau_ can be related to Sandawe _//ʔeḳa_, but it is hardly possible to prove that). However, there is a certain, if at first sight unnoticeable, match between NK and #Hoan. The glottalized uvular efflux is lost in PNK, like in II.93; the vocalic correspondences are very similar to the ones in II.72 (syllabic nasal developing to _-i-_ in #Hoan). The match with PKK and PNKK is actually harder to demonstrate; both the click influx and efflux are the same, but the correspondence PNK _*-ŋ-_ = PCK _*-ao-_ is not met anywhere else in this exact form. Cf., however, PNK _*šŋ̃_ "to be fat" (Žu. _šìᵑ_, !O. _čŋ̀_, etc.) and PCK _*cau_ (//Ani _cáú_, Naro _cau_, etc.) id.; it is possible that the _-o-/-u-_ element in CK derives from an earlier detachable class marker. With some caution, we may suggest a match here.

II.10. "BONE":

A) PNK __*!ʔú__ (//Au., !O. _!ʔu_, Žu. _!ʔú_); B) PT __*#âᵑ__ (!Xóõ _#âᵑ_, Mas. _//a_, |Nu//en _#aᵑ_); PKw __*!(o)a__ (? __*#(o)a__) (|Xam _!wa_, Batwa _!wa_); C) #Hoan __ćạa__; D) (?) PKK __*#xo__ (Nama _#kho-b_); E) PNKK __*|ʔóáᵑ__ (almost all languages have _|ʔóáᵑ_, exc. Deti, Kua _|ʔòáᵑ_, Tsua _|ʔóàᵑ_, Hie. _|ŋwa_); F) Sandawe __|í__; G) Hadza __mịƛa__.

Notes. The PT and PKw forms are the same, considering the frequent tendency to transcribe the palatal click as the alveolar one in !Wi data. Apart from that, no forms present any clear matches - not surprising considering the high "mobility" of the word for "bone" in Khoisan (thus, PKK _*#xo_ is hypothetical, since in !Ora the same root means "fruit kernel", while the actual word for "bone" is unknown; PNKK _*|ʔóáᵑ_, on the other hand, is comparable with Žu. _|ʔóáᵑ_ "leg", etc.).

II.11. "BREAST (chest)":

A) PNK __*!gòʔá__ (//Au. _!gwa_, Žu. _!gòʔá_); PT __*//gú__ (!Xóõ _//gúu_); PNKK __*//gǔ__ (//Ani, Naro, |Xaise, Deti, Cara, Tsixa, Danisi, Kua _//gǔ_, Tsua _//ǔ_, //Gana _//gû_); B) PKw __*//no[e]ᵑ__ (|Xam _//noain-tu_, //Ng _//nwoeŋ_, //Ku//e _//goin-tu_); C) #Hoan __!gạma__; D) PKK __*//xai__ (Nama _//khai-b_); E) Sandawe __saka__; F) Hadza __ʔiribá-__.

Notes. Case (A) is the first one in the wordlist where we actually meet with a possible correspondence of one click influx in one language branch to a different click influx in other branches - namely, NK _!_ vs. SK and CK _//_. As can be easily demonstrated, this is a rather frequent correspondence. Cf., for instance: PNK _*!ʔoaᵑ_ "to open" - PT _*//óʔa_ id.; PNK _*!òmà_ "short" - PT _*//ôhʔm_ "light, short, insubstantial in weight" - PCK _*//óm̀_ "short"; PNK _*!gùʔúbú_ "to swell" - PT _*//ûʰʔbu_ id.; PNK _*!gà̭_ "to belch" - PT _*//gàʰa_ id.; PNK _*!xò_ "to be unlucky" - PT _*//xō_ id.; PNK _*!gxòó_ "to be pregnant" - PT _*//gxoV_ id.; PNK _*!na̭ʔì_ "crowned plover" - PT _*//ṋâ̭eᵑ_ id., etc. To be precise, it must be noted that some of these PNK forms are postulated on the basis of Žu. evidence alone, and so may actually contain a PNK retroflex click instead of the alveolar one; however, such a correspondence (PNK _*!̭_ - PT, PCK _*//_) also exists, as will be demonstrated below.

II.12. "BURN (tr.)":

A) PNK __*kùʔú__ (//Au. _kou_, !Kung (Doke) _kuʔu_, Žu. _kùʔú_); B) PT __*Θʔá__ (!Xóõ _Θʔáa_, Mas. _ϑpwa_ "to make a fire"); C) PT __*//āʰa__ (!Xóõ _//āʰa_ "to set alight, singe", Mas., |Nu//en _//a_ "to burn"); PKw __*//a__ (|Xam, //Ng, //Ku//e _//a_, Bat., |Auni _//a_ "to cook"); D) #Hoan __Θui__; E) PKK __*dao__ (Nama, !Ora _dao_); PNKK __*dao__ (//Ani, Buga, Tsua _dáò_, Kxoe, Naro _dàó_, |Ganda, Deti, Cara, Danisi _dàò_, |Gwi, #Haba _díò_, //Gana, |Xaise, Tsixa, Kua _dáó_, Hie. _dhau_); F) Sandawe __kama__; G) Hadza __muli-__.

Notes. Proto-Taa has two different roots. **//a*, functioning as the main root for "burn" in Masarwa and |Nu//en, is obviously related to PKw **//a*; however, a match between PT **Θʔá* and #Hoan *Θuí* is hardly possible, not so much because of the vocalism but rather because there is no evidence for the correspondence "PT glottal stop efflux - #Hoan zero efflux" (not to mention that - running ahead - #Hoan and SK labial clicks practically never correspond to each other).

II.13. "CLAW (fingernail)":

A) PNK ***!̣ùʔúru*** (//Au. *//kuru*, Žu. *!ùʔúrú*, !O. *//kulu*); PT ****//qû[rV]*** (!Xóõ *//qûle*, pl. *//qûn-sâ*); PKw ****//urV*** (|Xam *//uru*, //Ng *//urisi*, #Kho. *//oro*, Bat. *//ola*, |Auni *//orasa*); (?) #Hoan ***!o̜ʔo***; PKK ****//oro*** (Nama *//goro-s*, !Ora *//òrò-b*); PNKK ****//oro*** (Naro *//oro* "finger, toe", |Gwi, //Gana (Tanaka)*!ore*, Deti *//óró*, Cara, Kua *//óròˋ*, Danisi *//áró*); B) PT ****//gàʔm*** (!Xóõ *//gà'm*); PNKK ****//à*** (//Ani, Tsixa, Danisi *//à*, Buga, Cara *//â*, |Xaise *//hâ*); C) Sandawe ***ç̣waʔa***; D) Hadza ***baλu***.

Notes. This is one of the most stable roots in the wordlist - found in all the main branches of Khoisan except for Sandawe and Hadza. The correspondences, however, need some explanation. For PNK **!̣* vs. #Hoan ! and SK/CK **//*, cf. II.54 ("moon"); PNK **!̣xom* "river" - PCK **//xom* id.; PNK **!̣nabu* "wing" - PCK **//abo* id. (Nama *//gawo-b*, etc.); some of the examples in II.11 can also belong here if the corresponding NK etyma have a retroflex click in PNK. For the uvular efflux in PT (and possibly PKw, since no recordings of !Wi languages mark the uvular consonants in that group), cf. PNK **!aʔò* "cheetah" - #Hoan *!a̜u^p* id. - PT **!qā^h* id.; PNK **!a̜è* "to hunt" - PT **!qāē^h* id.; whether the glottal stop in PNK **!̣ùʔúru* and the pharyngealization in #Hoan *!o̜ʔo* have anything to do with PT *-q-* remains to be established.

The #Hoan form is surprising in its lack of an inlaut resonant; it is possible that the matching is wrong, but considering the stability of the root almost everywhere, it is more likely that #Hoan either preserves the suffix-less root or features a peculiar development, e.g. **!o̜ʔro* > *!o̜ʔo*, with the resonant lost after a glottal stop.

There is also a possible second root, **//a-* or **//ga-*, found in PT and PNKK; the difference between the two is not quite clear, but some meaning nuances in !Xóõ suggest it could have been an early opposition between "a person's fingernail" and "hoof".

II.14. "CLOUD".

There is no special root for "cloud" in any of the major branches of Khoisan; it is usually denoted by composite forms like Žu. *!gà-!kxúí* or !Xóõ *!qhàa |Ghūa* (both literally meaning "rain-hair"). It is interesting to notice that the same root composition (see below the entries for "hair", "rain", and "water") is used in both NK and SK, yet for safety reasons we will not propose any root matches here. Occasionally we meet isolated obscure forms like Nama *!âu-s*, etc.; also Sandawe has ***ƛ̣uŋgu*** and Hadza has ***malundi-***, but none of these forms have any reliable Khoisan etymologies.

II.15. "COLD":

A) PNK ****#àʔú*** (!Kung *#áo*, Žu. *#àʔú*); PT ****//âʔu*** (!Xóõ *#âʔu^p*, (?) Mas. */kxau*, |Nu//en *//kʔau^p*); (?) #Hoan ***#a̜***; B) #Hoan ***/aba***; C) PKK ****!xai*** (Nama *!khai*, !Ora *!xai*); PNKK ****!qhai*** (Naro *!xaì̈*, //Gana *qai͡*, Hie. *haii*); D) Sandawe ***chawa***; E) (?) Hadza ***//naʎ̣a*** [uncertain].

Notes. The most important thing about this item is the match between NK and PT - the Žu. and !Xóõ forms, in particular, coincide fully but for the click influx. The resemblance is hardly coincidental:

the correspondence PNK _*#_ - PT _*//_ is confirmed by multiple examples, such as: PNK _*#ga_^{h} "old (of things)" - PT _*//à_^{hŋ} "old, mature"; PNK _*#kxóbó_ "to trample" - PT _*//xuBV_ id.; PNK _*#àù_ "giraffe" - PT _*//qhũ_^{ʔ} id.; PNK _*#o̭è_ "young" - PT _*//qṷV_ "new"; PNK _*#nà̭m_ "frog" - PT _*ʔ//na̭_^{h}_m_ id.; PNK _*#nŋ_ "Acacia fleckii" - PT _*//nà_^{hŋ} "Acacia hebeclada", etc.; cf. also II.80 from the wordlist. It is, however, unclear what is the corresponding click in CK languages, as no reliable matches in either PKK or PNKK have been found for any of these examples.

It should be noted that the PNKK form is reconstructed by me as _*!qhai_, with a uvular aspirated efflux, based on the data of //Gana (uvular articulation) and Naro (aspiration). Vossen's reconstruction does not include uvular consonants/click effluxes for PCK, and he is more inclined to treat occasionally met CK uvulars as innovation [Vossen 1992], but since the exact reason for such an innovation has not yet been determined, I find it reasonable to mark the presence of uvular reflexation in daughter languages by postulating uvular phonemes in the protolanguage, at least hypothetically.

II.16. "COME":

A) PNK **_*cí_** (//Au. _ci_, Žu. _cí_, !O. _ci_, _či_); PT **_*si/*sV_** (!Xóõ _sĩi_, _sa-_, Mas. _se_, _sí_; |Nu//en _sa_, _se_, _si_); PKw **_*sV_** (|Xam _sa_, _se_; //Ng, Bat., |Auni _sa_, _se_, _si_; #Kho., //Ku//e _sa_, _si_); Sandawe **_ʃi_**; B) #Hoan **_ča_**; C) PKK **_*ha_** (Nama _*ha_, !Ora _hà_); PNKK **_*hǎ_** (//Ani, #Haba, |Xaise, Cara, Danisi _hǎ_, Kxoe _jǎ_, Buga, Naro, |Gwi, //Gana, Tsixa, Deti, Kua, Tsua _hâ_, Hie. _ja_); D) Hadza **_ʒá-_**.

Notes. The NK and SK forms match perfectly (the correspondence of NK _*c_ to SK _*s_ is quite frequent, as demonstrated in [Honken 1988] and [Honken 1998]). The match with Sandawe is suggested on the same grounds as in II.7. #Hoan _ča_, however, is much more probable to belong together with PNK _*ča_ (Žu. _čá_ "to go and fetch"), despite the same root structure (affricate+vowel). Whether PCK _*ha_ and Hadza _ʒá-_ have anything to do with the NK/SK forms remains to be established; at present, there is no solid evidence to confirm a possible relationship.

II.17. "DIE":

A) PNK **_*!̣a̋i_** (//Au. _#ei_, !Kung _!̣ai_ (Dk.), Žu. _!a̋i_, !O. _//e_); #Hoan **_ši_**^{ʔ}; B) PT **_*|ʔâ_** (!Xóõ _|ʔâa_, Mas., |Nu//en _|ʔa_); PKw **_*|ʔa_** (nearly all languages have _|ʔa_); C) PKK **_*//ʔo_** (Nama _//o_, !Ora _//ʔo_); PNKK **_*//ʔó_** (all languages have _//ʔó_, exc. Tsua _ʔǒ_, Hie. _oo_); D) Sandawe **_ƛasi_**; E) Hadza **_tan//ʔi-_**.

Notes. The comparison of PNK _*!̣a̋i_ and #Hoan _ši_^{ʔ} may seem dubious at first. However, a careful investigation of the data shows that there is indeed a very probable correlation between several lexical items containing a retroflex click in NK and an initial hushing fricative (voiced or voiceless) in #Hoan. This includes such cases as "hand" (II.37); #Hoan _šiu_ "to dig" - Žu. _!gau_ (? < PNK _*!̣gau_) id.; and "water" (II.94). Additional #Hoan data (only about a dozen etyma with initial _š-_ or _ž-_ have been published) would help clear this connection further.

Many other sub-branches display some kind of lateral articulation in the root (CK and Hadza lateral click, Sandawe lateral affricate), but currently there is no significant evidence to relate PCK _*//(ʔ)_ to Sandawe _ƛ_; likewise, while the correspondence PNK _*!̣_ - PCK _*//_ does exist (see II.13), there are no examples of the PNK zero efflux corresponding to an "unwarranted" glottal stop efflux in PCK, so that we cannot propose a match between (A) and (C).

II.18. "DOG":

A) PNK **_*#ghŰ_** ((?) //Au. _!o_, !Kung _#ʔhwi_ (Dk.), Žu. _#ghű_^{ʔ}, _#ghőa̋_, !O. _#we_); PKw **_*#(ʔ)u_**^{ʔ}

(|Xam, //Ng *!ui*^{ŋ}, #Kho. *#ʔVn*, //Kxau *#huni*, Bat. *//ui*); (B) PT ***#qhà[-i]*** (!Xóõ *#qhài*, Mas. *#xai*, |Nu//en *#khi*); C) #Hoan ***cé[a]ma***; D) PKK ***ari*** (Nama *ari-*, !Ora *ʔàrì-b*); PNKK ***arV*** (//Ani *ʔérì-kù*, |Gwi *hárúgu*, #Haba *hǎrúgù*); E) PNKK ***ábà*** (Kxoe *ʔápà*, Buga, |Ganda *ʔápà*, //Gana, |Xaise, Cara, Kua, Tsua *ʔábà*, Deti, Tsixa, Danisi *ʔábá*, Hie. *aba*); F) Sandawe ***kāka***; G) Hadza ***//áʔano-***.

Notes. !Xóõ has two several roots for dog, one "main" (*#qhà*) and one specifically meaning "hunting dog" (*#gxū*). Their phonetic resemblance (palatal click + "complex" velar/uvular efflux) makes it hard to determine the exact relationship between these two roots and other subgroups' data. Yet the most probable solution is that PNK **#ghÚ-* and PKw **#(ʔ)ui*^{ŋ} (reconstruction very uncertain) are actually related to !Xóõ *#gxū*, both because of the vocalism (although this is not a decisive argument) and the voiced articulation in both PNK and PT (the exact efflux in PKw is, as in most cases, impossible to determine).

Outside of the NK/SK areal, however, there are no matches. #Hoan *cé[a]ma* (< **tema*) has no etymology, while the CK forms may be either expressive in origin, or old borrowings from Bantu (cf. proto-Bantu **-bua-* dog).

II.19. "DRINK":

A) PNK ****čhì(***^{ŋ}***)*** (//Au. *či*, Žu. *čhì*, !O. *či*^{ŋ}); PT ****kxā***^{hŋ} (!Xóõ *kxā*^{hŋ}, Mas. *kxa*^{ŋ}, |Nu//en *kxa*); PKw ****kx(o)a***^{ŋ} (|Xam *kxwa*^{ŋ}, *kxoe*^{ŋ}, //Ng *kxa*, *kxa*^{ŋ}, #Kho. *kxwa*^{ŋ}, Bat., |Auni *kxa*); #Hoan ***ču***; PKK ****kxa*** (Nama *a*, !Ora *kxâ*); PNKK ****kxâ*** (all languages have *kxâ* or, in case of East Central Khoe, *kʔâ*); B) Sandawe ***c̣ē***; C) Hadza ***fá-***.

Notes. The SK and CK forms are clearly related acc. to principle (a). Less obvious is their connection with the affricate-containing roots in NK and #Hoan. Cf., however, the following supporting evidence: PT **kxā*, PCK **kxě* "to cry, sound" - PNK **čʔí*^{ŋ} id.; PT **kxái*, PCK **kxâi*^{ŋ} "to laugh" - PNK **si~*chì* id.; and particularly "liver" (II.48). The variation between the affricates in NK (**čʔ*, **čh*, **ch*) is certainly questionable, but since a proper reconstruction of the PNK affricate/fricative system is still a task for the future, this cannot be a sufficient argument for rejecting the comparisons. Of special interest is the contrast between SK and CK, on one hand, agreeing on many PK items with initial *kx-*, and the near-total lack of good parallels with initial *kx-* in NK; it is very probable that the initial **kx-* was palatalized in all contexts, while NK *kx-* itself originates from different sources.

This treatment makes it impossible to track Sandawe *c̣ē* to the same source as the NK root - if PNK **čhì*^{ŋ} < **kxV*^{ŋ}, the Sandawe root obviously does not belong in the same group of etyma and has to be considered isolated.

II.20. "DRY":

A) PNK ****!̣kxau*** (!Kung (Doke) *!̣ʔau*, Žu. *!kxáú*); B) PT ****/ʔò*** (!Xóõ */ʔòo*); PKw ****/(ʔ)o*** (|Xam */owa*, Bat. */owa* "thirsty"); #Hoan ***/qʔau***, PKK ****/ʔo*** (Nama */o* "to become dry (of cow)", !Ora */ʔò*); PNKK ****/ʔô*** (Naro, Kua, Tsua */ʔô*); C) PT ****//ú̠*** (!Xóõ *//ú̠a*); PKw ****//o*** (|Xam *//o*); D) PNKK ****//xó*** (//Ani, |Gwi, //Gana, #Haba, Tsixa, Danisi *//xó*, Buga, Deti *//xǒ*); E) Sandawe ***/nʔ-***.

Notes. There is a solid isogloss here between SK and CK (**/ʔo*), with #Hoan */qʔau* very likely belonging here, acc. to the same correspondence pattern as in II.9 and II.93. On the other hand, while there is a certain level of resemblance between PNK **!̣kxau*, PT **//ú̠*, and PCK **//xo*, there is no evidence to support such a correspondence between click effluxes. Žu. does have both *//xó* "dryness (of the ground)" and */ʔò* "to be dry", not found in any other North Khoisan dialects; whether these are genetically related to the corresponding CK forms or are both old borrowings from CK cannot be determined at this point.

II.21. "EAR":

A) PNK ***/hűi*** (//Au., !O. */wi*, Žu. */hűi*); PT ***#nùʰ*** (!Xóõ *#nùhaᵑ*, Mas. *!nwa*, |Nu//en *#nu-ša*, pl. *#nu-i-te*); PKw ***#nu*** (|Xam *!nu-intu*, //Ng *!nwe*, *!nwe-ntu*, #Kho., |Au. *#nu-i*, |Nusan *!nu-du*); #Hoan ***/qhoe***, B) PKK ***#ae*** (Nama *#gae-b*); PNKK ***#é*** (//Ani, Kxoe, |Ganda, Naro, #Haba, Tsixa *#é*, Buga, |Gwi, //Gana *#ê*, |Xaise, Deti, Cara, Danisi, Cua *c̋e*, Kua, Tsua *kjê*, Hie. *čee*); C) Sandawe ***keke***, D) Hadza ***ha-š?apiči-***.

Notes. The match between PNK */hűi* and #Hoan */qhoe* is obvious (cf. for another example PNK **/húᵑ* "steenbok" - #Hoan */qhoᵑ* id.). The match between PNK and the SK forms, however, is only possible if the *-n-* efflux in SK is secondary - for instance, due to assimilation with subsequent nasal elements, such as the suffix **-ntu* in PKw or *-aᵑ* in !Xóõ. Such cases are indeed met rather frequently, if without an obvious regular pattern (cf., for instance, Žu. *!à'àmì* "to be in a circle" - PT **!ná?m* "to go round"; #Hoan *!ha̰na* "to snore" - PT **!na̰ʰna* id.). As for the correspondence PNK **/-* PT **#*, cf. the following examples: PNK **/n̥uì*, #Hoan */no̰e* "mouse" - PT **#nṵ̂*, PCK **#nú-ní* id.; PNK **/aboʰ* "to pile up" - PT **#Gàbo* id.; PNK **/a̰rì* "Acacia tortilis" - PT **#Ga̰ʰli* "Acacia fleckii"; PNK **/o̰* "bile, gall" - PT **#gáu* id.; PNK **/a̰?è* "to hold under the arm" - PT **#Gā̰ʰ?m* id.; PNK **/no̰ǐ* "to drown" (metathese from **/o̰ni*) - PT **#q?òni* id.; PNK **/nù?úᵑ* "to choke" - PT **#Gṵ́ʰnu* id., etc.

The CK forms thus match the NK/SK forms as far as the click influx, but the efflux (zero) and vocalism are crucially different, so no match can be postulated. Sandawe *keke* can theoretically be compared with PCK **#e* (< **#ké* < **cké* < **céke* < **keke*, acc. to the same syllable reduction principle as in II.41), but the phonetic development is too complex to be taken for granted without supporting evidence.

II.22. "EARTH":

A) PNK ***kxà*** (//Au., !O. *kxa*, Žu. *kxà*); #Hoan ***kxa***, B) PT ***#kxûm*** (!Xóõ *#kxûm*, Mas. *!um*, *//um*, *//kxom*, |Nu//en *!om-sa*); C) PKw ***!?au (*#?-)*** (|Xam, //Ng //Kxau *!?auᵑ*, (?) Bat. *//wa-lo* "ground", |Nusan *!?ouᵑ*); D) PKK ***!hu*** (Nama, !Ora *!hu-b*); E) PNKK ***xőm*** (//Ani, Kxoe, Buga, |Ganda, Naro, //Gana, #Haba, |Xaise, Cara, Tsixa, Danisi, Kua, Tsua *xőm*, |Gwi *xőám*, Deti *xòm*, Hie. *hom* "sand"); E) Sandawe ***!?uma***, F) Hadza ***jámu-***.

Notes. Another clear match between #Hoan and NK; outside of this isogloss, no definite parallels. Even the PKw form, while theoretically comparable with PT, cannot be fully understood (no trace of the velar affricate efflux anywhere). Likewise, a comparison of PKK **!hu* with Sandawe *!?uma* would require more examples of the PKK **-h-* - Sandawe *-?-* correspondence, which are absent.

II.23. "EAT":

A) PNK ***?m*** (//Au. *m̄*, Žu. *?m*, !O. *m*); PT ***?âᵑ*** (!Xóõ *?âᵑ*, Mas., |Nu//en *aᵑ*); PKw ***aᵑ~*eᵑ*** (|Xam, #Kho. *aᵑ*, //Ng, Bat. *aᵑ*, *eᵑ*); #Hoan ***?am***, B) PKK ***#?uᵑ*** (Nama *#û*, !Ora *#?uᵑ*); PNKK ***#?úᵑ*** (//Ani, Kxoe, |Ganda, Naro, //Gana, #Haba *#?úᵑ*, Buga, Tsixa *#?ǔᵑ*, |Xaise, Deti, Cara, Danisi, Kua, Tsua *?jǔᵑ*, Hie. *ŋjoo*); C) Sandawe ***mancha***, D) Hadza ***séme-***.

Notes. The same root is present in PNK and #Hoan; SK **?aᵑ~*?eᵑ* does not display the same consonant structure, but in case **?aᵑ* < **?ŋ*, this case is remarkably similar to the opposition between NK and SK 1sg personal pronoun (II.42), thus providing a solid match. The *-m-* vs. *-ŋ-* opposition, however, is still an important phonological isogloss between NK and #Hoan, separating them from SK.

II.24. "EGG":

A) PNK ***!nù*** (//Au., !O. *!nu*, Žu. *!nù*); B) PT ***#gú*** (!Xóõ *#gú-aᵑ*, Mas. *//w-a*, |Nu//en *!gu-oᵑ*);

PKw ***#(g)u** (|Xam *!aui*, #Kho. *#gw-i* "ostrich egg", |Auni *!u-ˀ* id.); C) #Hoan **čxui**; D) PKK ***!ʔubu** (***!kx-**) (Nama *!uwu-s*); E) PNKK ***#ʔúbí** (//Ani, Kxoe, Buga, Naro, |Gwi, //Gana, #Haba *#ʔúbí*, Deti *ʔjúbì*, Cara *ʔjìbí*, Tsixa *#ʔùbí*, Kua, Tsua *ʔíbí*, Hie. *ibi*); F) Sandawe **diʔa**; G) Hadza **ʔuλe-**.

Notes. Considering that PT distinguishes between several roots similar in both semantics and phonetics (apart from **#gú*, there is also **#úᵑ* "empty ostrich egg" and *#gūʰ* "sterile (of ostrich eggs)"), it is possible that some of the forms assigned to PKw belong to a different root, as there would practically be no possible way to distinguish between all these roots given the inferior quality of !Wi data transcription. It also makes any attempts at matching forms (A), (B), (D) and (E) extremely risky, even if some of them do look similar. Besides, there are serious phonetic problems here: Khoekhoe **!ʔ-* can hardly correspond to Non-Khoekhoe **#ʔ-*; and while some examples confirm that the correspondence PNK **!* - PT **#* is potentially valid (cf. PT **#xū-a* "elephant" - PNK **!xó* id.; PT **#ŋún* "navel" - PNK **!nÙʔm* id.; PT **#ʔàn* "penis" - PNK **!aʔm* id., with glottal stop metathesis in one of the forms), this leaves totally unexplained the nasal efflux in PNK; as mentioned in II.21, secondary nasal articulation in the efflux is practically always motivated in some way.

II.25. "EYE":

A) PNK ***|gàʔá** (//Au., !O. *|ga*, Žu. *|gàʔá*); #Hoan **Ɵoa**; B) PT ***!ʔû[-ʔ]** (!Xóõ *!ʔûᵑ*); C) PKw ***cʔaxu** (|Xam *cʔaxau*, //Ng *caxu*, #Kho. *cʔax(a)u*, //Kxau *cʔaxo*, Bat. *caxu, cau*, |Auni *cʔaxu*); D) PKK ***muᵑ** (Nama *mû-s*, !Ora *muᵑ-b*); E) PNKK ***#xa̋i** (//Ani, Buga, |Ganda, //Gana, #Haba *#xa̋i*, Kxoe, Naro, |Gwi *#xe̋i*, |Xaise *cа̋i*, Deti, Kua, Tsua *cxài*, Cara, Danisi, Cua *cxa̋i*, Hie. *čaii*); F) Sandawe **|we**; G) Hadza **ʔakhwa-**.

Notes. PNK **|* is one of the two possible correspondences for #Hoan *Ɵ*; for more examples cf. PNK **|ʔa̋u* "duiker" - #Hoan *Ɵʔu* id.; PNK **|naʔàᵑ* "sky" - #Hoan *ʔƟnoa* id.; and "head" (II.38). The efflux correspondence was spoken of earlier in II.4. Apart from this match, nothing definite can be suggested. PKK **muᵑ* is a secondary derivation from PCK **muᵑ* "to see" (see II.72); PNKK **#xa̋i* should be compared with PNK **#xa̋i* "to wake up", esp. since the PNKK root also has the same additional meaning in many languages. The PKw root is very unusual due to its peculiar bisyllabic structure (is *-xu* related in any way to PKw **xu* "face"?), but that in itself does not shed any light on its origin.

II.26. "FAT (n.)":

A) PNK ***|na̋i** (//Au., !O. *|ni*, Žu. *|na̋i*); B) PT ***sa̰ᵑ** (!Xóõ *sa̰ᵑ*, Mas. *šaᵑ*); PKw **so(e)ᵑ** (|Xam, #Kho. *soeᵑ*, //Ng *soa*); Sandawe **chaᵑ**; Hadza **hi-ça-**; C) #Hoan **|(ʔ)ui**; D) PKK ***//nui** (Nama, !Ora *//nui-b*); PNKK ***//nûi** (//Ani, #Haba, |Xaise *//ᵑnûi*, Kxoe, Buga, |Ganda, Naro, |Gwi, //Gana, Deti, Cara, Tsixa, Danisi, Kua *//ᵑûi*, Tsua *//ᵑùí*).

Root (B) is one of the most frequently used illustrations of the genetic relationship between all Khoisan sub-branches; to the PT, PKw, Sandawe, and Hadza forms must be added PNK **šŋ̃*, #Hoan *čaᵑ*, and possibly PCK **cau* "(to be) fat (adj./vb.)". For a more detailed account of the possible affricate correspondences in Khoisan see [Honken 1988]. It is possible that #Hoan *|(ʔ)ui* corresponds to PNK **|na̋i*, given at least several other examples of "secondary" arisement of #Hoan *-u-* (cf., for instance, PNK **ǯʔa*, PT **cʔà*, PCK **cʔaᵑ* "to steal" - #Hoan *ki-cuʔu* id.); however, the word is only found in an uncertain transcription in [Traill 1973], and until the click efflux is established properly, no final decision can be made.

II.27. "FEATHER".

Practically none of the major Khoisan subgroups have a root for "feather" that would be separate from the root for "hair", with the exception of PKK ***!ʔam** (Nama *!ammi*, !Ora *!amma*) and Sandawe ***thawa***; possibly also Hadza ***ha-i*** "feather, wing", found in [Bleek 1956]. None of these three roots have any evident connections; the rest will be discussed below in II.36.

II.28. "FIRE":

A) PNK ***dàʔá** (//Au., !O. *da*, Žu. *dàʔá*); B) PT ***/ʔà** (!Xóõ */ʔà^{ŋ}*, Mas., |Nu//en */a^{ŋ}*); PKw ***/ʔi** (|Xam, //Ng, #Kho., //Kxau, |Auni, |Nusan */ʔi*, Bat. */e*, */i*); PKK ***/ʔae** (Nama */ae-s*, !Ora */ʔae-b*); PNKK ***/ʔé** (//Ani, Buga, Cara, Tsixa, Danisi */ʔě*, Kxoe, |Ganda, Naro, |Gwi, #Haba, Kua, Tsua */ʔé*, //Gana */ʔê*, Hie. */ʔe*); C) #Hoan ***ϴgoa***; D) Sandawe ***//ʔʔ***; E) Hadza ***ço̩ko̩-***.

Notes. The match between SK and CK is obvious (the original vocalism is most probably **-e-*, with regular diphthongization in PKK and a regular **-e-> -a-* in PT, where *-e-*vocalism is extremely rare). PNK **dàʔá* may be related to PCK **dao* "burn" (see II.12), serving as a replacement for the original root. #Hoan *ϴgoa* is unclear.

II.29. "FISH":

A) PNK ***//ʔàù** (Žu. *//ʔàù*, !O. *//ʔau*); PCK ***//ʔau** (Nama *//au-b*, !Ora *//ʔàù-b*); PNKK ***//ʔâù** (//Ani, Buga, |Ganda, Naro, |Gwi, //Gana, #Haba, Tsixa *//ʔâù*, Kxoe *//ʔêù*, |Xaise, Deti, Cara *ʔáú*, Danisi *ʔâù*); B) Hadza ***/ʔama***.

Notes. Although PNK and PCK belong together according to criterion (a), this is one of the cases where a direct borrowing from CK is very probable. The word for "fish" is not generally widespread in either NK or SK (for SK, no reconstruction is given because the word is not available for *any* of the major SK languages, including !Xóõ!); even in Sandawe, *somba* "fish" is a Bantu borrowing (Proto-Bantu **-cómbà*). The only thing that still makes me - for now - count this as a possible match is the presence of *//ʔau* in at least two out of three clusters of NK dialects; however, a direct borrowing into PNK from PCK is not excluded either.

II.30. "FLY (vb.)":

A) PNK ***!no[-m]** (//Au., !O. *!no-a*, Žu. *!nom*); B) PT ***ʒā̰~*ʒo̰e** (!Xóõ *ʒā̰hi^{ŋ}*, Mas. *ʒoi^{ŋ}*, *ʒwe̩*); #Hoan ***zo̰e***; C) PKw ***//au** (|Xam *//au*, //Ng *//ou*); D) (?) Hadza ***ƛe̩ƛe*** [Bleek 1956].

Notes. A special root for "fly" is not reconstructible for PKK or PNKK (in most cases, the meaning is expressed by a root originally meaning "run", "move", "flee", etc.). Despite the general 'weakness' of the item, it still manages to yield a good isogloss between PT and #Hoan.

II.31. "FOOT":

A) PNK ***/kxa̋i** (//Au. */e*, */xe*, Žu. */kxa̋i*, !O. */kxe*); B) PT ***#nù** (!Xóõ *#nù^{ŋ}*, Mas. *#no*, |Nu//en *#nu*); PKw ***#n(o)a** (|Xam *!noa*, //Ng *//na*); C) #Hoan ***!gaʔu***; D) PKK ***#ʔai** (Nama *#ai-b*, !Ora *#ʔai-b*); E) PNKK ***!n̥are** (//Ani *!^{ŋ}nàré*, Kxoe *káré*, Buga, Tsua *káré*, |Ganda, Kua *kárì*, Naro *!nárè*, Cua *kàré*, Hie. *karee*); F) PNKK ***ʒî^{ŋ}** (Buga, |Xaise, Deti, Tsixa, Danisi *ʒî^{ŋ}*, Cara *ʒí^{ŋ}*); G) Sandawe ***//hata***; H) Hadza ***ʔapukwa-***.

Notes. The root seems to be very unstable. PNK **/kxa̋i* is possibly related to PSK **/kxa* 'hand' (see II.37), assuming the original meaning 'limb'; no better etymology can be provided. PKK **#ʔai* goes back to a PCK root with the meaning 'to kick'; the areal root **ʒî^{ŋ}*, meaning 'foot' in one major subgroup of East

Central Khoisan, probably has the original meaning 'toe' (cf. Nama _tsî-s_ "big toe"). As it is, no matches can be found between subgroups.

II.32. "FULL":

A) PNK __*!gàʔíᵑ__ (//Au. _!geᵑ_, Žu. _!gàʔíᵑ_); B) PT __*/ʔōrV__ (!Xóõ _/ʔōla_); C) PKw __*#auᵑ__ (|Xam _!auᵑ-_, //Ng _!xVŋ_, |Auni _//auᵑ_ "to fill"); D) PKK __*/kxoa__ (Nama _/oa_, !Ora _/kxőa_); PNKK __*/kxőe__ (//Ani, Buga, |Ganda _/kxôɛ̀_, Kxoe, |Gwi _/kxőɛ_, Naro, Danisi _/kxőe_, //Gana _/kxòè_, #Haba _/kxôè_, |Xaise, Kua _/ʔôè_, Deti, Cara, Tsua _/ʔőe_, Tsixa _/ʔôɛ̀_, Hie. _/we-ha_); E) Sandawe __lō-__; F) Hadza (?) __*//ʔnoso-__.

Notes. None of the forms (apart from PKK and PNKK, of course) seem to match. Even assuming that PT _*/ʔōrV_ < _*/ʔō_ with a secondary suffix, the correspondence between PT efflux _*-ʔ-_ and PCK efflux _*-kx-_ is not supported by any outside data.

II.33. "GIVE":

A) PNK __*/ʔàᵑ__ (//Au. _/a(ᵑ)_, Žu. _/ʔàᵑ_, !O. _/a_); PT __*/nV__ (!Xóõ _/nàa_ "dative formative", |Nu//en _/ni_ "give"); PKw __*/na__ (|Xam, //Ng _/na_, |Auni _/na_, _/no_); B) PT __*!qhá__ (!Xóõ _!qhá-_, Mas. _!xe_, _!xa_); #Hoan __šu__; C) PKK __*ma__ (Nama _ma_, !Ora _ma_); PNKK __*mâ__ (Kxoe, Naro, //Gana _mâᵑ_, |Gwi, #Haba _mâ_, Hie. _maa_); D) Sandawe __ie__; E) Hadza __kw(a)-__.

Notes. Of the two roots for "give" in PT, _*/nV_ is easily comparable with PNK _*/ʔàᵑ_ (either secondary nasalization in PT because of the nasal vowel or dissimilation in PNK). As for PT _*!qhá_ vs. #Hoan _šu_, cf. the strikingly similar case in II.94; apparently this correspondence fits in the same category as II.17, i.e. #Hoan fricative _š-/ž-_ vs. retroflex/alveolar click in PNK/PT. The vocalism discrepancy is not a big problem considering that PT _*!qhá_ can be < _*!qhó_ (another hypothesis is some kind of old ablaut as found in other verbal roots).

II.34. "GOOD":

A) PNK __*!àì__ (/Au. _!aiᵑ_, Žu. _!àì-sí_); PKw __*!eiᵑ__ (|Xam _!eiᵑ_, //Ng _!aiᵑ-ja_); PKK __*!aiᵑ__ (Nama _!gâi_, !Ora _!aiᵑ_); PNKK __*!âiᵑ__ (|Ganda, |Xaise, Deti, Cara, Tsixa, Danisi, Kua, Tsua _kâiᵑ_, Naro _!âiᵑ_, Hie. _kaie_ "agreeable, nice, pretty"); B) PNK __*žàᵑ__ (!Kung _ǯa_ (Doke), Žu. _žàᵑ_); C) PT __*qáiᵑ__ (!Xóõ _qáiᵑ_); PKw __*$t_1$ai__ (|Xam _twai_, //Ng _kiai_, Seroa _tae_, |Nusan _toai_); #Hoan __qhaen__; PNKK __*tʔÚ[i]ǹ__ (//Ani, Buga, //Gana, |Xaise, Tsixa, Danisi, Kua, Tsua _tʔûiᵑ_, Kxoe _tʔóñ_, |Ganda _tʔôǹ_, Naro _tʔúéᵑ_, Deti _tʔűiᵑ_); D) Sandawe __λā__; E) Hadza __cʔiʔeʔe-na__.

Notes. All the major subgroups have at least two roots for "good", with next to no differentiation in meaning. The first of these, the root _*!ai(ᵑ)_ is very well traced in NK, !Wi, and CK, without any serious phonetic disagreements between the groups.

The PKw root _*$t_1$ai_ is clearly related to PT _*qáiᵑ_; cf. such a similar case as !Xóõ _qhûje_ "ostrich" vs. |Xam _toe_, //Ng _kue_, #Kho. _twe_, etc.; note also that out of all the !Wi languages, //Ng shows a stable reflex _k-_ in both cases as opposed to the rest. The PKw phoneme _*t_1_ is postulated (as opposed to a regular _*t_ > _t-_ everywhere) to account for this correspondence. The exact character of this phoneme, as well as the mechanism of its development into _q-/k-_, becomes more clear when we compare it with PNKK _*tʔÚ[i]ǹ_: a glottalized _tʔ-_ can easily shift into a velar/postvelar position through partial assimilation with the glottal stop. So, while the match is not perfect, it is postulated on the following grounds: a) phonetic closeness of PT and #Hoan; b) confirmed correspondence between PT and PKw; c) phonetic closeness of PKw and PNKK.

II.35. "GREEN":

A) PNK **_*/a(u)ᵑ_** (//Au. _/gauᵑ_, Žu. _/auʰᵑ_, !O. _/aᵑ_); PT **_*/gā̰_** (!Xóõ _/gā̰hi_, Mas. _/ga̰i_); PKw **_*/aᵑ_** (|Xam _/ain_, _/ain-ja_); B) #Hoan **_zaʔa_**; Sandawe **_ʒaŋga_**; C) PKK **_*!kxam_** (Nama _!am_, !Ora _!kxam_); D) Hadza (?) **_//naweçě_**.

Notes. For the correspondence "PNK zero efflux - PT voiced efflux" cf. also PNK _*!aì̈_ "mortar" - PT _*!gâ[i]_ id.; PNK _*//àᵑ_ "to beg for" - PT _*//gàᵑ_ id.; PNK _*//à̰mà~*//à̰bà_ "to wear" - PT _*//gā̰ʰBV_ "to tie onto the body (skin, blanket)"; PNK _*//a̰ʔù_ "Ehretia rigida sp." - PT _*//gá̰u_ id., etc. More difficult is the possible connection between NK/SK forms, on one hand, and #Hoan _zaʔa_, on the other. There is at least one other possible case where #Hoan _z-_ is descended from a click: _zana_ "chin", cf., PNK _*!̰gaíʰᵑ_, PCK _*!gàn[í]_ id. However, the PT parallel there is _*ʒàni_, while the root for "green" also has a click reflex in PT; the two cases are thus far from being the same (not to mention that the latter one concerns a retroflex/alveolar click, whereas the former only concerns a dental one). For the moment, then, until we have more data on #Hoan, we will have to consider the two forms unrelated. #Hoan _zaʔa_ is, however, very similar phonetically to Sandawe _ʒaŋga_, and we can hypothetically match these two on the basis of rule (a).

For PNKK we do not have enough data to even suggest a possible proto-form, atlhough Naro _cǎ̰_ "blue, green" certainly suggests a possible affinity with the #Hoan-Sandawe isogloss (the other Naro form, _!am_ "dark green" is an obvious borrowing from Nama _!am_).

II.36. "HAIR":

A) PNK **_*!kxű̋i_** (//Au. _!kxwe_, Žu. _!kxű́í_, !O. _!wi_); PT **_*/Ghu_** (!Xóõ _/Ghùaᵑ_, Mas. _/wa-ni_); PKw **_*/khu_** (|Xam, //Ng _/u_, _/khu_, #Kho. _/khu[ke]_, Bat., |Auni _/kho_); PKK **_*/ʔuᵑ_** (Nama _/û-b_, !Ora _/ʔuᵑ-b_); PNKK **_*/ʔûᵑ_** (all languages have _/ʔûᵑ_ exc. for Cara _/ʔúᵑ_, Tsixa _/ʔùᵑ_, Hie. _/hoo_); B) #Hoan **_#nu_**; C) Sandawe **_çe_**; D) Sandawe **_!ʔu_**; D) Hadza **_há-ʎ̣è_**.

Notes. Despite the lack of immediate similarity between NK, SK, and CK forms, this is one clear example of how the assumption of "more-than-one-to-one" correspondences can shed additional light on Khoisan etymology. The most questionable correspondence here is PNK _*!_ - PSK _*/_, but there exists sufficient evidence to confirm it (and, as usual, where additional data exists, #Hoan sides with PNK, while PCK sides with PSK). Cf. the following examples: PNK _*!na̰ì̈_ "lion" - #Hoan _!haʔe_ id., but PT _*/á̰_ id.; PNK _*!gaíʰ_ "wildebeest" - #Hoan _!g(a)i_ id., but PT _*/à̰_ id., PCK _*/é_ id.; PNK _*!gú_ "belly, stomach" - #Hoan _!o_ id., but PT _*/ǒʰ_ id.; PNK _*!oʰm_ "dew" - PT _*/ùʰ_ id.; PNK _*!ùʔí_ "constipated" - PT _*/gūʔ_ id.; PNK _*!gxárú_ "to gnaw" - PT _*/gxùʔrV_ id.; PNK _*!gùʔúᵑ_ "to look" - PT _*/ṵ́_ id.; PNK _*!nòó_ "to long for, desire" - PT _*/nó_ "to desire intensely", etc.; cf. also "heart" (II.40), "name" (II.57), "red" (II.66), and "stone" (II.81) - altogether five examples in the 100-wordlist (and 4 of them belonging to the ultra-stable 35-wordlist part).

More complex is the problem of the click efflux. Normally, PNK _*-kx-_ does not correspond to PT _*-Gh-_ or PCK _*-ʔ-_ (the exact phonological nature of the PKw efflux is, of course, impossible to determine). However, the efflux _*-Gh-_ itself is quite rare in PT, and its exact correspondences in other families are yet to be determined; besides, !Xóõ itself yields a dialect variant _/qhùaᵑ_, so we cannot even be sure of the proper PT reconstruction. Elsewhere, PNK _*-i_ is easily explained as an old fossilized class suffix (_*!kxu-i_); cf. the same _*-i_ in "head" (II.38).

II.37. "HAND":

A) PNK **_*!̰gáú_** (//Au. _!gau_, !Kung _!̰gau_ (Doke), Žu. _!gáú_, !O. _//gau_); #Hoan **_šiu_**; PNKK **_*chàú_** (//Ani, Buga, Cara, Danisi _chàú_, Kxoe _çèú_, |Ganda, Naro, |Xaise, Deti, Tsixa, Kua, Tsua _càú_, |Gwi, //Gana *cáú*, #Haba *cáù*, Hie. *cau*); B) PT ***/kxà*** (!Xóõ */kxàa*, Mas., |Nu//en */kxa*); PKw ***/kxa*** (|Xam, //Ng, #Kho., //Ku//e, Bat., |Auni */kxa*, |Nusan */a*); C) PKK ***!Um*** (!Ora *!ùm̀-ma*; possibly Nama *!om-mi*, but the latter form irregular - **!gommi* should be expected); D) Sandawe ***ʞuᵑ***; E) Hadza ***ʔukhwa***.

Notes. The correspondence between the NK retroflex click and #Hoan *š-/ž-* has already been discussed in II.17. We would probably expect #Hoan **žiu* in view of the voiced efflux in PNK, but numerous examples given above seem to show that in many cases the voiced efflux may be secondary, due to reasons yet to be established. PNKK has **chàú* for hand, counted as a match with NK and #Hoan because of the striking similarity with the reflexation of "water" (II.96); add to this PCK **chàó* "to dig" - Žu. *!gàú* (? < **!gau*) id. - #Hoan *šiu* id., and the probability of chance resemblance moves close to zero.

II.38. "HEAD":

A) PNK ***/ŋa̋í*** (//Au., !O. */ne*, Žu. */ŋáí*); PT ***/nà*** (!Xóõ */nàn*, Mas. */na*, |Nu//en */nVŋ*); PKw ***/na*** (all languages have */na* or */naᵑ*); #Hoan ***ʔΘmuᵑ***; B) PKK ***dana*** (Nama *dana-s, dana-b*); C) PNKK ***mâ*** (//Gana, Deti, Cara *má*, |Xaise, Cua, Kua *mâ*, Danisi *mà*, Hie. *hma*); D) PNKK ***#ú*** (//Ani, Kxoe, Buga, |Ganda, Naro *#ú*; Tsixa *#ú*, *cúú*); E) Sandawe ***cẽ̄***; F) Hadza ***ʞóma***.

Notes. The match between PNK and PSK is obvious. For the match with #Hoan, see II.25, where the correspondence "PNK **/* - #Hoan *Θ*" has been demonstrated in more detail; note that the exact same correspondence, down to the click efflux, is also found in PNK **/naʔàᵑ* "sky" - #Hoan *ʔΘnoa* id. #Hoan *-u-* is probably secondary, as most of the #Hoan words beginning with a labial click contain a labial vowel.

None of the CK forms show any firm matches with NK/SK/#Hoan. It would be excellent if one could demonstrate a shift from **Θn* to **m* in PNKK, thus bringing PNKK **mâ* into comparison, but, unfortunately, it would be the only example of its kind, given the extreme rarity of initial **m-* in CKH. (However, cf. also PCK **mâ* "give" - PSK **/nV* id.? No #Hoan form, though).

II.39. "HEAR":

A) PNK ***càʔá*** (//Au. *ca*, *čʔa*, Žu. *càʔá*, !O. *sa*); #Hoan ***ca***; B) PT ***tá̰*** (!Xóõ *tá̰ᵑ*, Mas. *taᵑ*, |Nu//en *taᵑŋ*); C) PKw ***tu*** (|Xam, //Ng, //Kxau, |Auni *tu*, #Kho. *tjhu*, Bat. *tui*); D) PKK ***//nauᵑ*** (Nama *//nâu*, !Ora *//nàuᵑ*); Hadza ***//náʔe-***; E) PNKK ***kÚḿ*** (//Ani, Kxoe, Buga, |Ganda, Deti, Cara *kóḿ*, Naro, //Gana, #Haba, Tsixa, Danisi, Kua *kúḿ*, |Gwi *kúáḿ*, Tsua *cóḿ*, Hie. *čom*); F) Sandawe ***kheʔe***.

Notes. Despite the close resemblance of PT and PKw, the former rather belongs together with PKw **taᵑ* "to feel" (|Xam *taᵑ*, etc.); it cannot be, however, excluded that both go back to a single old stem with some kind of lexicalized ablaut. This question deserves a separate study; for now, we do not postulate a match here. Likewise, it would be tempting to put together PNK **càʔá* and PT **tá̰*, but we do not have additional supporting data to confirm this hypothetical affricativization of **t-* in PNK. On the other hand, we have a very interesting match between PKK **//nauᵑ* and Hadza *//náʔe-*.

II.40. "HEART":

A) PNK ***!kxá*** (//Au. *!a*, Žu. *!kxá*, !O. *kxa*); PT ***/qʔà[n]*** (!Xóõ */qʔàn*, Mas. */i*, |Nu//en */gaŋ*); PKw (?) ***/ɛ*** (|Xam */iᵑ*, //Ng */gai*, */ge*, #Kho. */e-kji*, //Kxau */ae*, //Ku//e */ɛᵑ*, |Auni */ɛ*, |Nusan */eᵑ*); #Hoan ***!qʔon***; B) PKK ***#ao*** (Nama *#gao-b*, !Ora *#áó-b*); PNKK ***#áó*** (//Ani, Kxoe, Buga, |Ganda, Naro, |Gwi, //Gana, #Haba, Tsixa *#áó*, |Xaise *cóóó*, Deti *cáò́*, Cara, Danisi *cáóó*, Cua *cóó*, Kua *kjó*, Tsua *kyô*, Hie. *čoo*); C) Sandawe ***ʒigida***; D) Hadza ***nkólo-***.

Notes. PT and PKw are easily grouped together, despite some discrepancies in the vocalism (the efflux _-qʔ-_ could not have even theoretically been marked in transcriptions of !Wi data). The #Hoan form matches PT according to the correspondence established in II.36 (PNK _*!-_ #Hoan _!-_ PSK _*/-_ PCK _*/_). The PNK form is more problematic; there is no additional data to support the correspondence PNK _*-kx-_ - PT _*-qʔ-_. On the other hand, there are no other firm examples of words with PT _*-qʔ-_ corresponding to anything in PNK (one such example, PT _*#qʔòni_ "to drown" - PNK _*/n̥ǒ́í_ id., has a secondary nasalisation in PNK due to the metathesis of the inlaut nasal, obscuring the original reflexation); and the possibility of the development _*-qʔ-_ > _*-kx-_ (considering that phonetically, _kx_ is an ejective affricate: _kx_ = _kxʔ_) is quite high. We can therefore suggest a temptative match here according to rule (d).

II.41. "HORN":

A) PNK __*!khú__ (//Au. _!u_, Žu. _!khú_, !O. _!khu_); #Hoan __!ho__; B) PT __*//a(e)ᵑ__ (!Xóõ _//āeᵑ_, Mas. _//Vn-ša_, |Nu//en _//aᵑ_); PKw __*//eiᵑ__ (|Xam, Bat. _//eᵑ_, //Ng _//aiᵑ_, #Kho., |Auni _//eiᵑ_); PKK __*//naᵑ__ (Nama _//nâ-b_); PNKK __*//nâ__ (all languages have _//ᵑnâ_); Sandawe __ƛana__; C) Hadza __loʔo-__.

Notes. The Sandawe/CK parallel (already present in [Ehret 1986]) is extremely interesting in that it shows one possible way for the secondary development of clicks in Khoisan out of consonant clusters formed through the reduction of original *CVCV > *CCV: _*//na_ < _*ƛna_ < _*ƛana_. This example is perhaps the most transparent of all; several more, although none of them with equally reliable phonology and semantics, can be found in [Ehret 1986]. To this root we might add PSK _*//aeᵑ_ (_*//eiᵑ_) assuming the possibility of a dissimilation in PSK (see the frequent fluctuations of the nasality in click effluxes discussed above). PNK and #Hoan show an isogloss of their own, in contrast to SK and CK.

II.42. "I":

A) PNK __*mV__ (//Au., !O. _m_, _me_, _mi_, Žu. _mí_); PT __*n̄__ (!Xóõ _n̄_, Mas. _n_, _na_, |Nu//en _ŋ_, _na_); PKw __*n/*ŋ__ (all languages have _n_ or _ŋ_); #Hoan __ma__; Hadza __(o)na__; B) PKK __*ti__ (Nama _ti[ra]_, !Ora _ti-re_ m., _ti-ta_ f.); PNKK __*ti__ (//Ani, Kxoe, Buga, |Ganda, #Haba _tí_, Naro _tíjá_, _tírá_, |Gwi _tírè_, //Gana _tê_, |Xaise, Deti, Cara, Danisi _tá_, Tsixa _tí_, Cua _tjé_, Kua _kjé_, Hie. _či_); Sandawe __ci__.

Notes. When it comes to the 1st person sg. pronoun, two types of forms are obviously distinguished - the form with a nasal (_*m-_, _*n-_) and the form with a dental/affricate (_*t-_, _*c-_). The question, then, is if they go back to no more than two roots or if this subgrouping does not reflect the actual historic situation.

It is possible that PNK/#Hoan _*m_ and SK _*n/*ŋ_ go back to different stems (both of them, interestingly enough, frequently found in different macrofamilies of the world). However, there is some significant evidence in favour of their being descended from one source. Comparing these phonological discrepancies with the ones found in the root "to eat" (II.23), we find the exact same correspondence - NK _m_ to SK _ŋ_. It is also necessary to notice that in many languages of the !Wi subgroup, the main pronoun _ŋ_ (or _n_) has a regular allomorph _m-_, _ma_ before words beginning with a labial consonant. Therefore we can suggest that the original form was _*ŋ_, with a later transition to _*m_ in PNK and #Hoan because of the loss of the 'unique' root structure (no other Khoisan root has a velar nasal as an individual phoneme, at least in the anlaut position), possibly triggered by assimilation in the pre-labial position. The Hadza form can be easily linked to the same source.

The alternate solution - i.e. treating PNK and PSK forms as descended from different roots - would look reasonable if there were at least some traces of both roots in both groups with a clear distinction of their possible functions (i.e. the opposition of direct/indirect stem, etc.). In SK, however, the variant with _m_ is, as indicated above, in complete phonological distribution with the _ŋ_-variant. In NK we do find certain traces of a *-n-* or *-ŋ-* in the function of 1sg personal pronoun. In Žu|'hoan, there is a special "dative" form *nà* "for me, to me", used postverbally or independently. Also, in [Bleek 1956] we occasionally find forms like *na* and *ŋ* either in the function of the subject ("I") or in the function of a possessive pronoun ("my") in several !Kung dialects. However, not a single dialect has any clear functional opposition of the two roots; it may well be that we are simply dealing with an archaic phonetic variant preserved in some places.

The CK/Sandawe parallel is likewise not a hundred percent convincing, as there is no regular correspondence "PCK **t-* - Sandawe *c-*". Yet there is nothing contradicting that correspondence, especially when it involves the sequence **ti-*, i.e. an environment in which the original dental is easily prone to palatalisation (which, by the way, does take place in some of the East Khoisan languages, such as Kua or Cua - irregularly, it should be noted). It is, in fact, an extremely important isogloss, and a very important argument in favour of the genetic relationship between Sandawe and the rest of Khoisan, as the morpheme **t-/*c-* for 1st person singular seems to be exclusive for that family, at least within African borders (much unlike **n*, **m*, or **ŋ*, all of which have a fairly wide distribution across the world).

II.43. "KILL":

A) PNK ***!khúᵑ*** (//Au., !O. *!uᵑ*, Žu. *!khúᵑ*); #Hoan ***lhoᵑ***; B) PT ***qâ[i]*** (!Xóõ *qâi*, Mas. *//kxai* "to kill by a blow on the head"); C) PKw ***/V*** (|Xam */a*, */i*, //Ng */a*, */i*, */khi*, #Kho. */kxa*); D) PKK ***!am*** (Nama *!gam*, !Ora *!am*); E) PNKK ***/kxúᵑ*** (//Ani, Danisi */kxǔᵑ*, Kxoe, Buga, Naro, |Gwi, //Gana, #Haba */kxúᵑ*, |Ganda */kxûᵑ*, |Xaise, Cara */ʔǔᵑ*, Tsixa */ʔùᵑ*, Kua, Tsua */ʔúᵑ*, Hie. */goo*); F) Sandawe ***ḳwe***; G) Hadza ***//ó-***.

Notes. The only obvious match here is between PNK and #Hoan. It would be tempting to join both of these forms with PNKK **/kxúᵑ*, seeing as how the correspondence "PNK **!* - PCK **/*" was established earlier in II.36. However, the same cannot be said about the correspondence "PNK **-kh-* - PCK **-kx-*", for which no further evidence can be found. In fact, PCK **/kxúᵑ* stands a better chance of finding a match in PKw **/V*, in case the #Khomani form */kxa* reflects a real *-kx-* efflux and the true PKw form is to be reconstructed as **/kx V*. It is, however, hardly reasonable to draw such an important conclusion on the basis of one dubious form; besides, there is still the vocalism discrepancy to be explained.

II.44. "KNEE":

A) PNK ***!γòà*** (//Au. *!wa-/ni*, Žu. *!γòà*); PT ***//γÚ*** (!Xóõ *//γúᵑ /nàn*, Mas. *//oᵑ-/naŋ*, |Nu//en *//guᵑ /nɪ*); PKK ***//oa*** (Nama *//goa-s*, !Ora *//óá-b*); PNKK ***//óé*** (//Ani *//óɛ́*, Deti *//òé*, Cara, Danisi, Tsua *//óé*, Tsixa *//úɛ́*, Kua *//üì*); B) PKw ***/nɔ*** (|Xam */no-aŋ*, */nu-aŋ*, //Ng */noᵑ*, Bat. */nu-ma*); C) #Hoan ***//ŋeme***; D) PNKK ***!úrù*** (Buga, |Ganda *kúrù*, Naro, #Haba *!úrù*, |Xaise *(kú)kúrù*, Cara *(kú)kúrù*, Hie. *kukuru*); Hadza ***guruŋguri-***; E) Sandawe ***ke(ᵑ)***.

Notes. Root (A) is characterized by an alveolar/lateral initial click (see more examples of this correspondence in II.11) and, according to both NK and SK evidence, a voiced velar fricative efflux. For more evidence on CK zero efflux corresponding to SK **-γ-* cf., for instance, PCK **//áùᵑ* "to fence" - PT **//γàuᵑ* "bush with hookthorns; to make a bush fence with hookthorns" - #Hoan *//xauᵑ* "fence" (there is no voiced *-γ-* in #Hoan). Curiously enough, both NK and SK demonstrate the same use of the root as first part of the composite **//γU-/n V*; the second part may very well be the same as the main PKw root for "knee" (**/nɔ-*), but the exact status of both roots and their relationship within Proto-Khoisan remains unclear.

Another interesting observation here would be to compare Hadza *guruŋguri-* with PNKK **!úrù*, especially those forms that show a reduplicated stem, like Hie. *kukuru*, etc. The exact phonological processes taking place here are hard to establish, but the resemblance is exceptionally striking, to the point of allowing us to postulate at least a temptative match.

II.45. "KNOW":

A) PNK ***!̣hàᵑ*** (//Au. *!haᵑ*, Žu. *!hàᵑ*); PNKK ***!ʔáᵑ*** (//Ani, Deti, Tsixa, Danisi *ʔǎᵑ*, Kxoe, Buga, |Ganda, |Xaise, Cara, Kua, Tsua *ʔáᵑ*, Naro, |Gwi *!ʔáᵑ*, //Gana *ʔâᵑ*, #Haba *kʔâᵑ*, Hie. *an*); B) PT ***/gûma*** (!Xóõ */gûmaᵑ*); C) PKw ***#ʔɛn*** (|Xam *#enn*, |Nusan *#an*); PKK ***#ʔan*** (Nama *#an*, !Ora *#ʔa̋n*); PNKK ***#ʔân*** (|Gwi, //Gana, #Haba *#ʔáǹ*, |Xaise, Deti, Danisi *ʔjàń*, Hie. *njin* "to think"); D) #Hoan ***ciʔa***; E) Sandawe ***mana***; F) Hadza ***çaha-***.

Notes. PNK **!̣* can apparently correspond not only to PCK *// (see II.13), but also to PCK **!*; apart from this example, cf. also PNK **!̣gaiʰᵑ* "chin" - PCK **!gàń[i]* id.; PNK **!̣gai̋* "puff-adder" - PCK **!gai* id.; PNK **!̣xàò* "hippopotamus" - PCK **!xáò* id. The correspondence PNK **-h-* - PCK **-ʔ-* is unsupported by additional data, but considering that there are no good matches in PCK for any other etyma with PNK **-h-*, we may suggest a temptative match according to rule (d).

The other widespread root is **#ʔan*, present in almost the exact same form in !Wi and CK. In CK it functions as the main root for "know" in Khoekhoe; the main meaning in PNKK seems to have been "think". On the other hand, **!ʔáᵑ*, the main PNKK root for "know", corresponds to Nama *!â* in *ho-!â* "to feel, perceive, learn", and probably to !Ora *!ʔaᵑ-b* "head".

II.46. "LEAF":

A) PNK ***dọ̀rà*** (//Au. *dora*, Žu. *dòạ̀rà*); #Hoan ***ʒ́ọba***; B) PT ***/gana*** (!Xóõ */gāna*, Mas. */gana*); PNKK ***/gàná*** (//Ani */gâᵑ*, Kxoe, |Ganda, Deti */gǎᵑ*, Naro, Hie. */gana*, //Gana */gánà*, |Xaise */gànà*, Cara, Tsixa, Danisi */gàná*); Sandawe ***/ā***; C) Hadza ***háçạ-phi***.

Notes. Since #Hoan *ʒ́* < **d*, we can easily compare its form with PNK, separating the elements **-ra* and **-ba* as different suffixes. Outside NK/#Hoan the more widespread root is **/gana*, found in the exact same form in PNKK and PT (in PKK **/gana* > **/aᵑ*, with the meaning shift "leaf" > "grass"; Nama */gâ-b*, !Ora */aᵑ-b*). A temptative match may also be suggested with Sandawe */ā* due to the same click influx; as for the voiced/voiceless character of the efflux, it fluctuates much too often between languages to form a serious objection.

II.47. "LIE":

A) PNK ***šú*** (//Au., !O. *šu*, Žu. *šú*); B) PT ***tû*** (!Xóõ *tûu*, Mas., |Nu//en *tu*); PKw ***tV*** (|Xam *ta*, *teᵑ*, *tiᵑ*, //Ng *tia*, //Kxau *ta*); C) #Hoan ***#qiʔi***; D) PKK ***//oe*** (Nama *//goe*, !Ora *//őe*); PNKK ***//őe*** (//Ani, Kxoe, Buga, Naro, //Gana, #Haba, |Xaise, Cara, Tsixa, Danisi, Kua *//őe*, Deti *//òé*, Tsua *//űe*); E) Sandawe ***//nine***; F) Hadza ***//angala-hi***.

Notes. No matches, apart from the obvious one between PKK and PNKK. PNK **šú* and PT **tû* do look suspiciously similar, but the **š* - **t* correspondence is not supported by any other convincing data (on the other hand, there are at least several good examples of correspondences between PNK **š* and PT **s*).

II.48. "LIVER":

A) PNK ***čhíᵑ*** (//Au. *či*, !O. *čiᵑ*, Žu. *čhíᵑ*); PKK ***kxaiᵑ*** (Nama *âi-b*, *âi-s*, !Ora *kxaiᵑ-b*); PNKK ***kxa̋iᵑ*** (//Ani, Buga *kxàíᵑ*, Kxoe *kxáíᵑ*, |Ganda *kxȁiᵑ*, Naro, Danisi *kxai̋ᵑ*, |Gwi, //Gana, #Haba, Tsixa *kʔai̋ᵑ*, |Xaise *kʔíᵑ*, Deti, Cara *kʔàíᵑ*, Cua *ćʔíᵑ*, Kua, Tsua *cʔǐᵑ*, Hie. *če*); B) PT ***//nàm*** (!Xòõ *//nàm*, |Nu//en *//n Vm*);

PKw ***//n(o)aᵑ*** (|Xam *//noaᵑ*, //Ng *//nain*, (?) //Kxau *ŋaŋa*); C) #Hoan ***kui***; D) Sandawe ***thas(i)no***; E) (?) Hadza ***//ʔneya-kho***.

Notes. The correspondence between PCK **kx-* and PNK **čh-* fully matches the one described in II.19 (curiously enough, the NK development here is fully parallel to the palatalization in certain East Central Khoisan languages). On the other hand, the resemblance between these forms and #Hoan *kui* is probably accidental, since, as seen in II.19, #Hoan should agree with PNK in this palatalisation process. It is theoretically possible that the palatalisation could be prevented in certain contexts, but until we have more data from both #Hoan and PNK/PCK, there is nothing else toillustrate such a suggestion.

The Hadza form (acc. to Derek Elderkin's data) is not compared with the SK forms, partially because of vocalism problems, partially because it is somewhat suspicious (it is not quite clear what is actually the main word for 'liver' in Hadza).

II.49. "LONG":

A) PNK ***#gàʔV́ᵑ*** (//Au. */geᵑ*, !Kung (Doke) *#gaʔŋ*, Žu. *#gàʔíᵑ*); B) PT ***!ʔa*** (!Xóõ *!ʔám* sg., *!ʔáᵑ* pl.); C) PKw ***/ʔa*** (//Ng, //Kxau */ʔa*); D) #Hoan ***čaʔa***; E) PKK ***ga[i]xu*** (Nama *ga[i]xu*, !Ora *gáxù*); F) PNKK ***!áò*** (//Ani, Buga, Naro, #Haba, Kua, Tsua *!áò*, |Gwi *!àù*, //Gana *!áù*, Deti, Tsixa, Danisi *káò*, Cara *káó*); G) Sandawe ***maganʒa***; H) Hadza ***thase-***.

Notes. All the branches and sub-branches seem to have a separate root for this word, frequently used also to denote "tall" and/or "deep".

II.50. "LOUSE":

A) PNK ***#ṇà*** (Žu. *#ṇà*); B) PNK ***cʔíᵑ*** (Žu. *cʔíᵑ*); #Hoan ***cʔi***; C) PT ***ϑnV̰́*** (!Xóõ *ϑnṵ́ᵑ*, pl. *ϑnàᵑ-tê*); PKw ***ΘnU*** (|Xam *Θnuiŋ*, //Ng *Θnoin-ja*); Sandawe ***māŋ/ʔa***; Hadza ***/ʔámaçi-***; D) PKK ***kxuri*** (Nama *uri-*, !Ora *kxúrĩ-b*); PNKK ***kxuni*** (//Ani *kxùní*, Buga, |Gwi *kxúní*, //Gana *kxúnì*, |Xaise, Deti, Cara, Kua, Tsua *kʔúnì*, Tsixa *kʔùní*).

Notes. The SK-Sandawe-Hadza comparison is another interesting example (see [Ehret 1986]) of the possible secondary development of clicks - in this case, the labial click in SK is explained through the influence of the original labial nasal. The Hadza form is thus the archaic variant, with PSK ** ΘnV* < **ΘmV* < **/mV* < **/ʔᵃmV*; the Sandawe form features a metathesis of the original structure. Of course, this is but a temptative match, but it perfectly agrees with the same model of development that was exposed in II.41.

The two PNK forms are very approximate, as Žu|'hoan is the only NK dialect for which the word "louse" is actually recorded, and there is no way of determining which form is the main one and which one is secondary. The form **#ṇá* bears a strong resemblance to PT **ΘnV̰́*, right down to the pharyngealized vowel, however, no other fully convincing examples of the correspondence "PT ** Θ* - PNK **#*" have been found (a much more probable correspondence is PNK **!̰*, see II.53). The other form, **cʔíᵑ*, is obviously the same as #Hoan *cʔi*.

In Žu|'hoan there is also a form *kxúrí* "sp. of louse", obviously tied in with PCK **kxu-[ri/-ni]* and less obviously with PT **//gxóni* "species of louse". Regardless of whether the Žu|'hoan form is a borrowing from CK or related to it genetically, it seems to have a more restricted meaning and so cannot be taken into account.

II.51. "MAN":

A) PNK ***!hóáᵑ*** (//Au. *!waᵑ*, Žu. *!hóáᵑ*, !O. *!uᵑ*); B) PT ***Tâ*** (!Xóõ *tâa*, Mas., |Nu//en **da*); C) #Hoan

ạri; D) PKK ***kxao** (Nama *ao-b*, !Ora *kxao*); PNKK ***kxáò** (//Ani, Kxoe, Buga, Naro, |Gwi, #Haba *kxáò*, //Gana *kxáó*, |Xaise, Cara, Tsixa, Kua, Tsua *kʔáò*, Deti *kʔáó*); E) Sandawe ***maxe***; F) Hadza ***λeme***.

Notes. The PKw root for "man" is impossible to determine (most languages have their own individual way of expression, see [Bleek 1929]). Likewise, PT does not have a special root and uses the word for "person" instead; NK, #Hoan, and CK distinguish "man" from "person", but none of the forms appear to be related.

II.52. "MANY":

A) PNK ***#khái̋** (//Au., !O. *#khi*, Žu. *#khái̋*); B) PT ***//á-** (!Xóõ *//áli*, Mas. *//ari*, |Nu//en *//ante*, *//arri*); C) #Hoan ***ʒua***; D) PKK ***#ui** (Nama *#gui*); E) Sandawe ***dẽ***, F) Hadza ***ndago***.

Notes. Neither PKw nor PNKK yield a good candidate for the root; apparently, it was hardly stable at the Proto-Khoisan level either. #Hoan *ʒua* has, however, an interesting parallel in !Xóõ *ʒâa* "multitude, crowd".

II.53. "MEAT":

A) PNK ***!̣khá** (//Au. *!a, !kha, //a*, Žu. *!khá*, !O. *//kha, //a*); PT ***Θà** (!Xóõ *Θàje*, Mas. *Θwe*, |Nu//en *Θwe*); PKw ***Θa** (|Xam, //Ng *Θwai*, #Kho. *Θoi*, Bat. *Θwa*, |Auni *Θwe*); B) #Hoan ***//ạe***, C) PKK ***kxo** (!Ora *kxo-b*); PNKK ***kxó** (//Ani, Danisi *kxǒ*, Kxoe, Buga, |Ganda, Naro, #Haba *kxó*, |Gwi, //Gana *kxô*, |Xaise, Deti, *kʔǒ*, Cara *kʔò*, Tsixa, Kua, Tsua *kʔó*, Hie. *koho*); D) Sandawe ***/nin***; E) Hadza ***mana-ko***.

Notes. Out of all the possible alternatives, the NK retroflex click looks like the most promising correspondence for SK labial clicks. Cf., apart from this example, the following: PNK **!̣ó* "elder brother" - PT **Θxà* id.; PNK **!̣háᵑ* "son, child" - PT **Θqâ̰* "child"; also, with either PNK alveolar or PNK retroflex, PNK **!gom* (**!̣-*) "edible caterpillar" - PT **Θgòᵑ* id.; PNK **!gòàᵑ* (**!̣-*) "Kalahari raisin bush" - PT **ΘGhùᵑ* id. More problematic is the aspiration in PNK, which makes the match less reliable; most of the time PNK *-kh- corresponds to PT **-h-* or **-qh-*. It is, however, possible that the SK root was contaminated with the ancestor of !Xóõ *Θàᵑ* "herd of eland, flesh, meat", and thus lost the original aspiration.

II.54. "MOON":

A) PNK ***!̣nűi** (//Au. *!nwi*, Žu. *!núí*, !O. *//nwi*); PNKK ***//nóé** (//Ani *//ᵑnóɛ́*, Kxoe, Buga, Tsixa *//ᵑóɛ́*, |Ganda, Naro, Deti, Cara, Danisi, Kua *//ᵑóé*, |Gwi, //Gana *//ᵑóè*, #Haba *//ᵑnóè*, |Xaise *//ᵑnóé*, Tsua *//ᵑúé*); B) PT ***!qhàn** (!Xóõ *!qhàn*, Mas. *!xVn*, |Nu//en *!xan*); C) PKw (?) ***#ʔoro** (|Xam *!a!auru*, //Ng *!orre*, #Kho. *#ʔɔrɔ*, //Kxau *#ʔoro*, //Ku//e *tʔɔlo*, Bat. *ƛolo*); D) PKK ***//xaᵑ** (Nama *//khâ-b*, !Ora *//xaᵑ-s*); E) Sandawe ***!a-biso***; F) Hadza ***sétha-***.

Notes. The best match here is between PNK and PCK; see II.13 ("claw") for more examples of the **!̣ - *//* correspondence. Whether PT **!qhàn* and PKK **//xaᵑ* have anything to do with each other still remains to be established. The PKw root is exceptionally interesting in that some of the languages, most notably Batwa (//Xegwi) have a lateral consonant instead of the click in the anlaut position; again, it is not yet clear whether this lateral should be reconstructed for PKw or PSK, and if so, what are its origins and correspondences in other branches.

II.55. "MOUNTAIN".

In most major subgroups of Khoisan the word for "mountain" is the same as the word for "stone", with the following exceptions: #Hoan has ***!hu*** (probably related to !Xóõ *!ùʰm* "hill, niche for trees");

Sandawe has __gawa__; and Hadza possibly has __//ʔuɬe__. None of these three words are related, and matches for the other languages will be discussed under "stone" (II.81).

II.56. "MOUTH":

A) PNK __*cʔí__ (//Au., !O. _ci_, Žu. _cʔí_); B) PT __*#û__ (!Xóõ _#û-e_, Mas. _!w-e_, |Nu//en _#u-eⁿ_); C) PKw __*tu__ (|Xam, //Ng, #Kho., //Ku//e, Bat., |Auni _tu_, |Nusan _du_); D) #Hoan __šʔ__; E) PKK __*kxam__ (Nama _am-s_, !Ora _kxam_ "gate"); PNKK __*kxáḿ__ (all West CK languages have _kxáḿ_; |Xaise, Cara, Tsixa, Cua _kʔáḿ_, Deti, Kua _kʔàḿ_, Tsua _kʔàm̀_, Hie. _#am_); F) Sandawe __!mum__; G) Hadza __ʔawanika-__.

Notes. Unless there are some deeply hidden, complex correspondences in this root, none of the forms seem to match. PNK _*cʔ-_ never seems to reflect a palatalised _*kx-_, and even if it did, there would still be the vocalism and final _-m_ to explain. Likewise, the resemblance of PNK and #Hoan is deceptive, since NK hissing consonants always correspond to #Hoan hissing ones (unless the transcription of the #Hoan word is actually incorrect).

II.57. "NAME":

A) PNK __*!ú__ (//Au., !O. _!u_, Žu. _!ú_); PT __*|à(u)ⁿ__ (!Xóõ _|àuⁿ_, Mas. _|kxauⁿ_, |Nu//en _|áⁿ_); PKw __*|e(ⁿ)__ (|Xam _|eⁿ_, _|e_, //Ng _|eⁿ_, |Auni _|e_, _|en_); #Hoan __!o__; PKK __*|kxon__ (Nama _|on-s_, !Ora _|kxònná_); PNKK __*|kxôn__ (//Ani, //Gana _|kxôn_, Kxoe _|kxón_, Buga, |Ganda, Naro _|kxûⁿ_, |Gwi _|kxòàn_, #Haba, Danisi _|kxûn_, |Xaise, Cara _|ʔôn_, Deti _|ʔúń_, Tsixa _|ʔòǹ_, Kua, Tsua _|ʔûn_, Hie. _|un_, _//un_); B) Sandawe __//wa__; C) Hadza __ʔakana-__.

Notes. The word for "name", oddly enough, is often among the most stable elements of the 100-wordlist, and Khoisan is no exception. PNK and #Hoan, as usual, are closer to each other phonetically. PT _*|à(u)ⁿ_ perfectly matches them according to the correspondence exposed in II.36. As for the CK forms, there indeed is evidence for CK _*-kx-_ sometimes corresponding to a zero efflux in NK/SK, cf., for instance, PKK _*|kxau_ "dew, spring" - PNK _*!oʰm_, PT _*|ùʰ_ "dew"; PKK _*#kxoni_ "worm" - PNK _*#nùʔúⁿ_ id. (secondary nasal efflux due to assimilation with the nasal vowel). PCK also shows some irregular fluctuations of the vocalism, including an unexplainable diphthongisation into either _-ui-_ or _-oa-_; taking into account the PKw vocalism _-e-_, one might suggest that the original PK form contained some kind of diphthong, possibly _*-eu-_, which later gave rise to all the untrivial vocalic developments.

Neither Sandawe nor Hadza, however, cannot be successfully linked to the PK root without making a handful of unverifiable assumptions.

II.58. "NECK":

A) PNK __*!ȁȉⁿ__ (//Au. _!eiⁿ_, Žu. _!ȁȉⁿ_, !O. _//áⁿ_); B) PT __*#kxàuⁿ__ (!Xóõ _#kxàuⁿ_, |Nu//en _#uⁿ_); PKw __*#ʔau__ (|Xam _!au_, _!khou_, //Ng _!u_, //Kxau _#ʔu_, |Auni _#oiⁿ_); C) #Hoan __ća__; D) PKK __*!kxao__ (Nama _!ao-b_, _!ao-s_, !Ora _!ʔáó-b_); PNKK __*!kxáó__ (//Ani, Naro, #Haba _!kxáó_, //Gana _kxáó_, |Xaise _kʔàò_, Deti, Kua _kʔàó_, Cara _kʔáó_, Cua _kʔáò_); E) Sandawe __ḳweg__; F) Hadza __|úti-ja__.

Notes. PT _*#kxauⁿ_ and PCK _*!kxao_ look extremely similar, but the correspondence "PT *# - PCK *!" is not found anywhere else; for the moment we should consider this as a mere chance resemblance.

II.59. "NEW":

A) PNK __*zé__ (//Au. _ze_, Žu. _zé_, !O. _ʒe_); #Hoan __za__; Hadza __ʒana__; B) PT __*//qụV__ (!Xóõ _//qụV_, Mas. _//xwe_); PKw __*//ue__ (|Xam _//we_); C) PKK __*kaba__ (Nama _kawa_); PNKK __*qábà__ (Naro _kạba_, |Gwi, //Gana _qábà_); D) Sandawe __//ae__.

Notes. A good match here between PNK and #Hoan, and we can furthermore add Hadza *ʒana* because of the phonetic resemblance. Sandawe and SK both have a lateral click efflux, but the resemblance ends there - and we would be expecting a bisyllabic root in Sandawe anyway, so as to account for the **//q-* anlaut in PSK.

II.60. "NIGHT":

A) PNK ***/gú*** (//Au., !O. */gu*, Žu. */gú*); B) PT ***/nų́*** (!Xóõ */nụ́e*, Mas. */nọ́e*, |Nu//en */noe*); C) PKw ***//ga*** (|Xam, //Ng, Bat., |Nusan *//ga*, #Kho. *//ʔa*, *//ga*, //Kxau *//a*, |Auni *//gau*, *//go*); D) #Hoan ***chao***; E) PKK ***thu*** (Nama *tsu-xu-b*); PNKK ***thǔ*** (//Ani, Kxoe, Buga, |Ganda, Cara, Tsixa, Danisi *thǔ*); Sandawe ***twē***; F) Hadza ***çifi-***.

Notes. Most of the roots are unrelated. PNK **/gú* and PT **/nụ́* are quite similar, but there is nothing to explain the nasality in PT (or its loss in PNK?); as we have seen many times earlier, every time a nasal influx "irregularly" appears where it should not be expected, it can be explained by the influence of an ensuing nasal element, not present in this root. On the other hand, PCK **thu* and Sandawe *twē* are quite similar and most probably belong together.

II.61. "NOSE":

A) PNK ***cʔú^ŋ*** (//Au. *čú^ŋ*, Žu. *cʔú^ŋ*, !O. *cuŋ*); B) PT ***/nù^h*** (!Xóõ */nùhná*, Mas. */nu*, */nu-ča*, |Nu//en */nuša*); PKw ***/nu*** (|Xam */nu(n)tu*, //Ng, #Kho., //Ku//e */nutu*, Bat. */nu*, |Auni */nu*, */no^ŋ*, |Nusan */nudu*); Sandawe ***/nathi***; Hadza ***ʔinthawe-***; C) #Hoan ***/qʔo^ŋ***; D) PKK ***#ui*** (Nama *#gui-s*, !Ora *#űi-b*); PNKK ***#uî*** (//Ani, Buga, |Ganda, |Gwi, Tsixa *#uî*, Naro, #Haba *#űi*, //Gana *#guî*, |Xaise, Danisi, Cua, Kua, Tsua *cúî*, Deti, Cara *cúi̋*, Hie. *čui*).

Notes. SK forms are grouped together with Hadza and Sandawe based on the presence of the combination "click+nasal" or "stop+nasal" in all three branches, quite typical for the word "nose" in macrofamilies all over the world. This last factor is the chief reason why it is reasonable to make an exception from the stricter rules described in the introduction; there are no other cases of Sandawe */n-* directly corresponding to Hadza *-nt(h)-* (cf., however, a somewhat similar case in II.80), or to PT **/n-*, yet there is nothing to disprove such a correspondence either, and the meaning "nose" is hardly coincidental for these structures. It is not excluded that PNK and #Hoan actually belong here as well (the former through secondary affricativisation of the click), but this is a very feeble hypothesis so far unsupported by further evidence.

II.62. "NOT":

A) PNK ***/ó́V́*** (//Au. */wa*, Žu. */óá*, !O. */wa*, */wi*, */we*); #Hoan ***/hoʔon***; B) PT ***//qhuV́*** (!Xóõ *//qhúa*, Mas. *//kʔa*, |Nu//en *//u*); PKw ***//V*** (|Xam *kxau* (?), //Ng *//u*, *//e*, *//au*, #Kho. *//e*, *//ai*, *//o*, *//ɯ*; C) PKK ***tama*** (Nama, !Ora *tama*); PNKK ***ta*** (Naro *-tá*, *-tàmà*, //Gana *tàmà*, |Xaise, Deti, Cara, Danisi, Kua, Tsua *-tà*, Tsixa *-ta^ŋ*); D) Sandawe ***-çe***; E) Hadza ***a-kwV*** (Bleek).

Notes. Each major subbranch of Khoisan seems to have its own primary negative morpheme, except for PNK and #Hoan, which form an obvious match. (#Hoan *-h-* is a little problematic, but words of this category - frequently used particles, etc. - may allow for slight irregularities; also the exact #Hoan form probably needs verifying). Their vocalism often varies due to assimilation with juxtaposed words.

II.63. "ONE":

A) PNK ***/nèʔé*** (//Au. */ne*, Žu. */nèʔé*, !O. */ne*, */nee*); #Hoan ***Θmuᵖ***; B) PT ***#ʔû*** (!Xóõ *#ʔu-*, Mas. *!uʔe*, |Nu//en *!oe*); PKw ***//ue~*!ue*** (|Xam *!wai*, //Ng *//we*, #Kho. *//oe*, Bat. *//a*, |Auni *#uᵖ*); C) PKK ***/ui*** (Nama */gui*, !Ora */ui*); PNKK ***/űí*** (all languages have */űí*); D) Sandawe ***çexe***, E) Hadza ***ʔičame-***.

Notes. #Hoan and PNK match according to the correspondence laid out in II.35 ("eye"); #Hoan vocalism is secondary here due to assimilation with the labial click. The SK forms are rather strange, displaying an untrivial variation of clicks rarely met elsewhere - such as the lateral click in //Ng, Batwa, and #Khomani vs. the alveolar click in |Xam and the palatal click in !Xóõ and |Auni. This is not the same variation as when we deal with a simple palatal click, "yielding" supposedly alveolar reflexes (or, rather, supposedly transcribed as alveolar) in |Xam (see, for instance, II.24), because here the #Khomani form, where the palatal click is regularly marked, contains a lateral click. Whether this points to a "fifth" click in PKw, like in PNK, is yet to be determined; for now the one thing that is certain is that the PKw root does seem to match the PT one.

II.64. "PERSON":

A) PNK ***žù*** (//Au. *žu*, Žu. *žù*, !O. *ǯu*, *žu*); PT ***tâ*** (!Xóõ *tâa*, pl. *tûu*, Mas., |Nu//en *da*); B) PKw ***!ui*** (|Xam, //Ng *!wi*, //Kxau *!wi* "man", Bat. *kwi* (?), |Nusan *!gu*, *!gui*); C) #Hoan ***#ʔam-koe***, D) PKK ***khoe*** (Nama, !Ora *khoe-*); PNKK ***khőé*** (//Ani, Buga, |Ganda, |Gwi, |Xaise, Deti, Cara *khőé*, Naro, //Gana, #Haba, Tsixa, Danisi *khóè*, Kua *khòé*, Tsua *chòé*, Hie. *čwa*, *čowe*); E) Sandawe ***/nome-***; F) Hadza ***ʔúnù-***.

Notes. Normally PNK **ž* can be shown to correspond to PT **ʒ* (cf. PNK **ža̯m* "thin" - PT **ʒá̯ba* id., PNK **žàbì* "to turn round" - PT **ʒá̯bi* id., etc.). However, there is some evidence showing that the correspondence "PNK **ž* - PT **t*" is also valid. Cf., besides "person" (where the discrepancy in vocalism is explained through an old, somewhat obscure ablaut, cf. !Xóõ plural *tûu*), PNK **žó* "black" - PT **tò*^{hŋ} "dark"; PNK **žom* "paw, fist" - PT **tà*^{h} "footpad"; PNK **ža̯ʔà* "blood" (rare form, as opposed to the commonly widespread **/ʔŋ*) - PT **tà-rV* "clotted blood". In this context it is interesting to note that the reconstruction **tâ*, with initial **t-*, is made dubious because of the presence of a dialectal form *la*, found in Masarwa [Bleek 1956]. This, of course, reverts us to the question of possible SK laterals, first raised in II.54. Lack of data from SK dialects, unfortunately, does not allow us to make any definite conclusions - but it is quite probable that PNK **ž* actually stems from two earlier phonemes, namely, **ǯ* (yielding PT **ʒ*) and either lateral **ƛ* or **λ* (yielding PT **t*).

In #Hoan the second element of the composite used to denote "person" is obviously related to PCK **khoe*, but it is the first element that should be used for comparison, and the root **#ʔam-* so far does not have any plausible etymology.

II.65. "RAIN":

A) PNK ***!gà*** (//Au. *!ga*, !Kung *//ga*, *!!ga* (Lloyd), Žu. *!gà*, !O. *//ga*); PKw ***!kh(o)a*** (|Xam *!khwa*, //Ng *!kha*, //Kxau *!ʔa*); B) PT ***!kxôe*** (!Xóõ *!kxôe*, Mas. *!we*, |Nu//en *!xwe*); C) #Hoan ***čoʔaᵖ***; D) PKK ***tu*** (Nama *tu-s*, !Ora *tú-s*); PNKK ***tú*** (all languages have *tú* exc. for Buga, Deti *tŭ*); Hadza ***ʔathi-***; E) Sandawe ***ƛ̣waᵖ***.

Notes. The words for "rain" and "water" in Khoisan seem to be often related; the most obvious situation is in Hadza, where there exists only one word for both notions. In PT there seem to be two different roots (**!kxôe* for "rain" and **!qhà* for "water", see below); however, PKw, at least, according to the poorly transcribed data in our possession, does not differentiate between the two either. In this respect it is interesting to compare the NK roots *!gà "rain" and *!gú "water", the only difference between which lies in the vocalism - which, furthermore, can be explained by the same kind of ancient ablaut that we see in "person" (II.64). If this is so, the *-a*-form of the root would have the primary meaning "rain", and then PNK and PKw are perfectly comparable; for more details on the problem, see II.96 ("water").

The Hadza form has been compared with PCK before, and although the vocalism correspondences are unclear, this does not prevent us from postulating a temptative match according to rule (a) (total or near total identity of consonant structure).

II.66. "RED":

A) PNK ***!gà^{ŋ}*** (//Au. *!ga^{ŋ}*, Žu. *!gàà^{ŋ}*, !O. *!gai*, *!ga^{ŋ}*); PT ***/ā^{h}-ná*** (!Xóõ */āhná*, Mas. */anja*); #Hoan ***/aʔa***; B) PKK ***/kxaba*** (Nama */awa*, !Ora */kxàbà*); C) PNKK ***/(n)òá*** (//Ani *ʔnòá*, Buga, Deti *ʔóá*, Naro, #Haba *ʔnòá*, |Gwi *ʔòà*, //Gana *ʔnóà*, |Xaise, Cara, Danisi */òá*); D) Sandawe ***buɬi***; E) Hadza ***tekise*** (Bleek), ***tese*** (Elderkin).

Notes. PNK matches with #Hoan and PT according to several correspondences already established above (for PNK **-g-* vs. #Hoan zero efflux, cf. II.4; for PNK and #Hoan **!* vs. PT **/*, cf. II.36). Whether the root can further be compared with PNKK **/(n)òá* is unclear, because there is still the nasality to be accounted for in PNKK, even if it appears and disappears somewhat sporadically. The PKw root for "red" cannot be established, as there are numerous forms in [Bleek 1929] and [Bleek 1956], with almost none of the languages agreeing with each other.

II.67. "ROAD/PATH":

A) PNK ***#khà*** (//Au., !O. (?) */a*, Žu. *#khà*); B) PT ***dào*** (!Xóõ *dào*, Mas. *dau*); #Hoan ***ʒéo***; PKK ***dao*** (Nama *dao-b*, !Ora *dàò-b*); PNKK ***dâò*** (//Ani, Buga, Naro, //Gana, Tsixa, Danisi *dáò*, Kxoe, |Xaise, Deti, Cara, Kua, Tsua *dáó*, |Ganda *dáù*, |Gwi, #Haba *díó*, Hie. *dhau*); C) Sandawe ***//ō***; D) Hadza ***jeke***.

Notes. The most widespread root for "road" in Khoisan is **dao* (no relation to modern Chinese!), present in all the branches except for PNK (and PKw, where again no proper reconstruction can be suggested). Unfortunately, there is no way so far to make sure it is not really a cultural borrowing from PKK into all the other families; however, for the time being we will make the assumption of genetic relationship between all these forms.

II.68. "ROOT":

A) PNK ***//àrì*** (Žu. *//àrì*, //Au. *//ari* "root fibre", !O. *//ale*, *//are* id.); B) PT ***!kxá*** (!Xóõ *!kxáì*); C) #Hoan ***!qʔai***; D) PKK ***!noma*** (Nama *!noma-b*, !Ora *!nòmà^{ŋ}-b*); E) Sandawe ***!nʔ***; F) Hadza ***wili*** (Elderkin).

Notes. A very poor, unstable item; language data shows that the word for "root" very frequently assumes a more localized meaning, e.g. "root of a particular plant" or "root fibre", and vice versa. For PKw and PNKK it is more or less impossible to suggest a protolanguage form, due to lack of data in general and/or lack of comparable forms. !Xóõ and #Hoan forms are similar, but the correspondence "PT **-kx-* - #Hoan *-qʔ-*) does not exist, and the resemblance appears to be a chance one.

II.69. "ROUND".

No major Khoisan branch has anything even remotely approaching a protolanguage root for "round"; the closest would be PKK ***!ʔuBV*** (Nama *!ubu*, !Ora *!um*), which does not have any external parallels anyway. We will have to exclude this root from our calculations.

II.70. "SAND".

In most Khoisan languages the word for "sand" is the same as the word for "earth", with no differentiation at all; thus, matchings for most branches are the exact same ones as for "earth" (II.22). Elderkin records ***ç̣awa*** "sand" for Hadza, as opposed to ***jamu-*** "sand", but this is not much help.

II.71. "SAY":

A) PNK ****kU*** (!Kung *kue* (Lloyd), Žu. *kò*); PKw ****kV*** (|Xam *ka*, #Kho. *ka*, *ku*, *kwa*, //Kxau *ku*, |Auni *ko*); B) PT ****tám*** (!Xóõ *tám*); C) PKK ****mĩ***^{ŋ} (Nama *mî*, !Ora *mí*^{ŋ}); PNKK ****mí(***^{ŋ}***)*** (//Ani, Kxoe, Naro *mí*^{ŋ}, #Haba, Cara *mí*, Hie. *me*); D) Sandawe ***bo***; E) Hadza ***he-***.

Notes. Curiously enough, none of the recorded forms have any clicks, probably reflecting the frequency of their use and the semi-particle status of many of them. Despite this, only PNK and PKw forms actually agree in having **kV* as the basic structure; PT prefers **tV* (cf. also !Xóõ *tâna* "to talk, speak"); and CK languages have the extremely rare **mV* structure.

II.72. "SEE":

A) PNK ****sŋ̃*** (//Au. *se*, !Kung *sn* (Doke), Žu. *se*, !O. *siŋ*, *sŋ*); #Hoan ***ci***; B) PT ****|nâ*** (!Xóõ *|nâ*^{ŋ}, Mas. *|na*, |Nu//en *|ne*); PKw ****|na~*|ne*** (|Xam *|na*, *|na*^{ŋ}, *|ne*, *|ni*, //Ng *|na*, *|ne*, *|ni*, #Kho. *|na*, *|ne*, *|ni*^{ŋ}, //Kxau, |Auni *|na*, //Ku//e *|ne*, Bat. *|na*, *|ne*); C) PKK ****mu***^{ŋ} (Nama *mû*, !Ora *mu*^{ŋ}); PNKK ****mû***^{ŋ} (//Ani, Kxoe, Buga, |Xaise, Deti, Cara, Tsixa, Danisi, Kua, Tsua *mû*^{ŋ}, Naro, #Haba *mô*, |Gwi, //Gana *m̂*, Hie. *moo*); D) Sandawe ***|a***; E) Hadza ***kli-***.

Notes. PNK **-ŋ̃* is used here to denote the "i-tinged" syllabic nasal (which in some dialects alternates with *-e-* and *-i-*), as opposed to the "a-tinged" syllabic nasal found, for example, in **|ʔŋ* "blood". The PNK form itself is obviously related to #Hoan *ci*, according to the same vocalic correspondence as found in "blood" (II.9); for the **s - c* correspondence, cf. also PNK **sì* "they" - #Hoan *cì* id., etc. No other matches are found, although it is interesting to note that PSK has **|n* while PCK has **m* - - just like in "head" (II.38) and "give" (II.33); Sandawe *|a* can also be viewed only as a very hypothetical match with PSK, since the effluxes do not match.

II.73. "SEED":

A) PNK ****//àʔá*** (Žu. *//àʔá*); B) PT ****sâʔ***^{ŋ} (!Xóõ *sâʔa*^{ŋ}); C) #Hoan ***!uru***; D) PNKK ****|xúrí*** (//Ani, |Gwi, //Gana, #Haba, Deti, Cara, Danisi, Kua, Tsua *|xúrí*, (?) Buga *|xùí*, Naro *|xùrí*, Hie. *|khuri*); E) Sandawe ***bojo***; F) Hadza ***|otu*** (Bleek).

Notes. Also a very weak root, only relatively stable in PNKK; for PKw and PKK it is unreconstructible, and even in PNK and PT not all dialects agree. Despite the huge variety of forms, none present any firm matches.

II.74. "SIT":

A) PNK ****|nŋ́*** (//Au. *|ni*, Žu. *|ńŋ*, !O. *|ŋ*); #Hoan ***ʔ|na***; PKK ****#nu***^{ŋ} (Nama *#nû*, !Ora *#*^{ŋ}*ú*); PNKK ****#nú***^{ŋ} (//Ani, Kxoe, #Haba *#*^{ŋ}*nú*^{ŋ}, Buga, |Ganda, //Gana *#*^{ŋ}*nú*, Naro *#*^{ŋ}*nú(*^{ŋ}*)*, |Xaise, Kua *ńú*^{ŋ}, Deti *nú*^{ŋ}, Cara, Tsixa, Danisi *ńú*, Tsua *ńù*^{ŋ}, Hie. *njo*); B) PT ****chû*** (!Xóõ *chû*, Mas. *ču*, |Nu//en *šu*, *ču*); PKw ****so(***^{ŋ}***)*** (|Xam, //Ng *so*, *soeŋ*, #Kho. *sou*, Bat. *šo*, |Auni *sa*^{ŋ}, *sao*^{ŋ}); C) Sandawe ***hakiç̣(i)***; D) Hadza ***hama***.

Notes. PNK does not distinguish between the preglottalized nasal efflux and the simple nasal efflux, which makes the match with #Hoan quite justified (cf. also PNK **|nòm* "springhare" - #Hoan *ʔ|nam* id., etc.). Moreover, there is a good match with PCK - the correspondence "PNK, #Hoan */- PSK, PCK *#" had previously been established in II.21 ("ear"), and the effluxes correspond to each other directly. The same root is thus lacking only in both SK branches which instead display a root with an initial hissing fricative/affricate (Masarwa and |Nu//en forms with hushing sounds are non-diagnostic since we also meet many cases where hissing and hushing reflexes appear to be in free variation), and in Hadza/Sandawe.

II.75. "SKIN":

A) PNK ***/nó*** (//Au., !O. */no*, Žu. */nó*); B) PT ***t(ʔ)ṵ̀m*** (!Xòõ *tṵ̀m*, Mas. *tʔym*, |Nu//en *tʔum*); PKw ***tu***^{ŋ} (|Xam *tu*^{ŋ}, //Ng *tu*^{ŋ}, *twa*^{ŋ}, #Kho. *gjo*); #Hoan ***čʔu***; C) PKK ***kho*** (Nama *kho-b*, !Ora *khō-b*); PNKK ***khǒ*** (//Ani, Buga, Tsixa, Danisi, Kua, Tsua *khǒ*, Kxoe *kxò*, |Ganda, Naro, |Xaise, Deti, Cara *khò*, |Gwi, //Gana, #Haba *khô*, Cua *chǒ*, Hie. *čo*); D) Sandawe ***kelemba***; E) Hadza ***ʔáhà-***.

Notes. #Hoan *čʔu* is counted as a match with PT and PKw for the following reasons: 1) all the forms bear strong phonetic resemblance; 2) no other correspondences for #Hoan *čʔ-* have been found so far; 3) there exists evidence showing that the initial phoneme in PSK wasn't merely a simple **t-*; Masarwa and |Nu//en data are recorded with a glottalised phoneme, and #Khomani shows a rare reflexation *gj-*, only found in a few other cases (*gjisi* "what" < **Tisi*, see "what") - so the glottalisation in #Hoan actually agrees with the SK data.

II.76. "SLEEP":

A) PNK ***cʔá*** (//Au., !O. *ca*, *cʔa*, Žu. *cʔá*); #Hoan ***ca***; B) PT ***Θâ[i]n*** (!Xóõ *Θân*, Mas., |Nu//en *Θwoin*); PKw ***ΘVn*** (|Xam *Θoeŋ*, //Ng *Θoeŋ*, *Θwoiŋ*, #Kho. *Θʔo*^{ŋ}, //Kxau *Θan*, Bat. *Θweni*, |Auni *Θwaʔi*^{ŋ}, |Nusan *Θuin*); C) PKK ***//ʔom*** (Nama *//om*, !Ora *//ʔùm*); PNKK ***//ʔóm̀*** (Naro, //Gana, #Haba *//ʔóm̀*, |Gwi *//ʔóam̀*, Danisi *//ʔùḿ*, Hie. *//gom*); Sandawe ***//ʔo***; D) Hadza ***//úpi-***.

Notes. PNK and #Hoan match exactly, apart from the unclear glottalisation in PNK; however, observe the same correspondence in "tooth" (II.89). Another exact match is found between PCK and Sandawe, a rare case of both click efflux and influx being the same. Hadza *//úpi-*, however, cannot be placed here for the time being, as the influxes do not match, and the bisyllabic character of the root requires additional explanation.

II.77. "SMALL":

A) PNK ***cʔèmà*** (//Au., !O. *cema*, Žu. *cʔèmà*); Sandawe ***c̣ō***; B) PT ***/ʔûi*** (!Xóõ */ʔûi*); #Hoan ***/kxui***; C) PKK ***/a*** (Nama */ga*, !Ora */à*); D) PNKK ***/kxáré*** (Naro, |Gwi, #Haba */kxáré*, Deti, Kua, Tsua */ʔàré*, Cara, Danisi */ʔáré*, Hie. *kare* "a little"); E) Hadza ***kụmi-***.

Notes. Sandawe and PNK forms feature the same consonant structure (PNK **-mà* is originally a diminutive suffix; the simple form *cʔe* is also found in a couple of dialects) and thus form a possible match. As for the other roots, some of them find parallels in other branches: PT **/ʔûi* and #Hoan */kxui* (for the efflux correspondence cf. II.8) can be compared with PNK **/ùʔí* "to be thin", while PKK **/a* finds a parallel in PT **/qa̰-* "diminutive formative for pronouns". However, only PT and #Hoan present an exact wordlist match. The PKw form for the root cannot be suggested.

II.78. "SMOKE":

A) PNK ***šórV̀*** (//Au. *šore*, *šori*, Žu. *šórà*, !O. *čule*, *čuli* "tobacco"); B) PT ***ckxâjV*** (!Xóõ *ckxâje*); PKK ***/kxan*** (Nama */annis*, !Ora */kxàn*); PNKK ***cʔán[ì]*** (//Ani, Kxoe, Buga, |Ganda *cʔánì*, Naro, //Gana,

#Haba, Tsixa, Danisi, Kua, Tsua *c?ínì*, |Gwi *c?ɛ́nɛ̀*, |Xaise, Deti, Cara *c?ání*, Hie. *cene*); Sandawe ***c̣uḳa***; Hadza ***c̣iḳo-wa***; C) #Hoan ***ʒ́u̯e***.

Notes. Root (b) is one of the most interesting cases in the wordlist; this comparison, already present in [Ehret 1986] and several other sources, is an exceptionally strong argument in favor of "macro-Khoisan" relationship. The basic root structure **c̣Vḳ V*, preserved in Hadza and Sandawe, develops into **c̣ḳV* through regular reduction of the first syllable; this structure is clearly seen in PT **ckxâ-jV* and PCK **c?a-ni* - the latter form could also serve as an indirect argument in favour of interpreting the correspondence "PKK **|kx* - PNKK **c?*" as the cluster **ckx* rather than the glottalised affricate **c?* in Vossen's reconstruction. There are, of course, no other firm examples of the same correspondence, but the supposed developments look perfectly logical and do not contradict any other established or hypothetical correspondences.

In PNK the original root was replaced by **šórV̀*, with unclear connections. #Hoan *ʒ́u̯e* (where *ʒ́* < **d*) is obviously connected with Žu. *do̯e* "to smoke out (bees), to inhale smoke".

II.79. "STAND":

A) PNK ****!núᵑ*** (//Au. *!nu*, Žu. *!núᵑ*, !O. *!nw-a*); PT ****//hûᵑ*** (!Xóõ *//hûᵑ*, Mas. *//uᵑ*, *//oᵑ*, *//huᵑ*, |Nu//en *//hu*, *#hu*); B) #Hoan ***!ui***; C) PKK ****mâᵑ*** (Nama *mâ*, !Ora *mâᵑ*); D) PNKK ****té*** (//Ani, Kxoe, Buga, |Ganda *tɛ́*, Naro *téᵑ*, |Gwi, Danisi, Kua *té*, //Gana *tê*, Deti *tě*, Cara, Tsua *tè*, Tsixa *tɛ̂*, Hie. *the*); E) Sandawe ***//nume***; F) Hadza ***?ika-***.

Notes. For the click influx correspondence between PNK and PT, see II.11; the irregularity between the effluxes is again explained by the influence of the nasal vowel in PNK, leading to the secondary nasalisation of the efflux. It is tempting to compare the forms with Sandawe *//nume*, but there are too many problems with this comparison - the nature of the second syllable in Sandawe, the question of whether the lateral or the alveolar manner of articulation is primary, and the question of the click influx (is assimilation with the following nasal element also characteristic of Sandawe? This question is practically impossible to answer for now).

II.80. "STAR":

A) PNK ****#uʰᵑ*** (//Au. *#goeᵑ*, Žu. *#uʰᵑ*, !O. *#uᵑ*); PT ****//ōna*** (!Xóõ *//ōna*, Mas. *//gwana-te* pl., |Nu//en *//ana-te* pl.); PKw ****//oV*** (//Ng *//kxwe-sa*, #Kho. *//?wai-kje*, //Ku//e *//an-te* pl.); #Hoan ***#oᵑ***; B) PKK ****|(n)amVro*** (Nama *|namiro-s*, !Ora *|ȧmȯrȯ-b*); C) PNKK ****|xánì*** (//Ani, |Xaise *|xání*, Buga, |Ganda, Deti, Tsixa, Danisi, Kua *|xánì*, Cara *|xínì*, Tsua *|xáinì*, Hie. *#khaine*); D) Sandawe ***|nowaᵑ***; Hadza ***ncha***.

Notes. NK, #Hoan, and SK forms all match according to the correspondence established in II.15; the efflux is zero in every case, and the nasalisation in PNK and #Hoan may reflect an earlier nasal consonant, still preserved in many SK languages. To these forms one should add Naro *#o̯nu*; although the root has a very scant distribution in CK, it can hardly be qualified as a borrowing from NK because of phonetic dissimilarities.

Sandawe *|nowaᵑ* and Hadza *ncha* are joined together due to the similarity between this case and II.61 ("nose"); in both cases, Sandawe *|n* corresponds to a Hadza initial cluster "*n* + dental/affricate". Whether these forms have anything to do with PKK **|(n)am Vro* remains to be seen.

II.81. "STONE":

A) PNK ****!nÙm*** (//Au., !O. *!num*, Žu. *!nòm*); PT ****|nū*** (!Xóõ *|nūle*, pl. *|nūn*, |Nu//en *|nyle* "stone", *|nun* "mountain"); B) PKw ****!au*** (|Xam *!au*, *!au-ken*, //Ng *!au*, //Kxau *!ao*, |Nusan *!gou*); C) #Hoan ***//ho̯a***; D)

PKK ***/ʔui** (Nama */ui-*, !Ora */ʔui*); PNKK ***//nóá** (//Ani *//^{ŋ}nóá*, Kxoe, Buga, |Ganda, //Gana, Tsixa, Danisi *//^{ŋ}óá*, Naro, #Haba *//^{ŋ}nọ̀ạ́*, |Gwi *//^{ŋ}òá*, Kua *//^{ŋ}óà*, Hie. *//gwa*); E) Sandawe ***diṅ***; F) Hadza ***ha!ʔá-***.

Notes. PNK and PT match according to the correspondence established in II.36 (provided that PNK **-m* is originally a detachable suffix). Apart from that, there are no matches, even if #Hoan *//họa* is comparable with PNKK **//nóá*; the nasal efflux in NKK, however, asks for additional explanation.

II.82. "SUN":

A) PNK ***/am** (//Au., !O. */Vm*, Žu. */ám*); PNKK ***/ám̀** (all languages have */ám̀*); B) PT ***//ʔân** (!Xóõ *//ʔân*, Mas., |Nu//en *//Vn*); PKw ***//ʔVn~*//ʔV^{ŋ}** (|Xam *//oi^{ŋ}*, //Ng *//oe^{ŋ}*, *//oi^{ŋ}*, #Kho. *//ʔui^{ŋ}*, //Kxau *//ʔoe*, Bat. *//oi^{ŋ}*, //un); Sandawe ***//ʔaka-su***; C) #Hoan ***čha***; D) PKK ***sore** (Nama *sore-s*, !Ora *sórė-b*); Hadza ***ʔišo-***.

Notes. PNK and PNKK present a perfect match (note that in PKK the same root is also found, but only in the meaning "to heat up, be hot" - Nama */gam* - which makes the idea of borrowing into NK from CK much less probable, since most of these borrowings are of Khoekhoe origin).

There are also interesting isoglosses between Sandawe and SK, on one hand, and PKK **sore* and Hadza *ʔi-šo-*, on the other hand. In the latter case PKK *-re-* is detachable as a derivative suffix, and Hadza *ʔi-* is one of those "classificatory" prefixes whose function and usage are still not quite clear due to lack of an extensive description of the grammatical/derivational structure. #Hoan *čha* is unclear, but possibly connected with words like Nama *tsē-b* "day", etc., all of Bantu origin.

II.83. "SWIM":

A) (?) PNK ***ǯxà** (Žu. *ǯxà*); B) (?) PKw ***ϑxu** (|Xam *ϑxu*); C) PNKK ***thá^{ŋ}** (Nama *tsâ*, !Ora *thá^{ŋ}*); D) PNKK ***bàrá** (//Ani, Kxoe, Buga, Naro, Deti, Cara, Tsixa, Danisi *bàrá*, |Xaise *bàrà*); Sandawe ***phuduse***; E) Hadza ***//oʔo-***.

Notes. Apart from CK, this item is very poorly recorded in other branches - absent even in Traill's extensive !Xóõ dictionary. If the PNKK root is to be transcribed as **bàdá* (**-d-* and **-r-* are in free variation in CK; I prefer to transcribe the consonant as **-r-* simply because it agrees better with the 'classic' scheme of the Khoisan disyllabic root as seen also in NK and SK material), it is easily comparable with Sandawe *phudu-se*. That said, one should note that initial **b-* is a very rare phoneme in PCK, and most CK words beginning with this consonant (or the voiceless *p-*) are Bantu borrowings, either recent or going back to the PCK stage itself; this makes the comparison somewhat unstable.

II.84. "TAIL":

A) PNK ***!̣xűi** (//Au. *!khwi*, Žu. *!xűi*, !O. *//we*); (?) PKw ***!khui** (|Xam *!khwi*); #Hoan ***Θxui***; B) PT ***/àu^{ŋ}** (!Xóõ */àu^{ŋ}*, Mas. */àu^{ŋ}*, */gau^{ŋ}*, |Nu//en */gau^{ŋ}*); C) PKK ***sao** (!Ora *sáó-b*, (?) Nama *sao* "to follow, go in single file"); PNKK ***cáó** (all languages have *cáó* except for Kxoe *tçáó*, #Haba *càó*, and Hie. *cau*); Sandawe ***çwa***; Hadza ***çáhó-***.

Notes. Root (C) is an important (and previously well-known) cognate not only between Hadza and Sandawe but also, as it seems, between them and CK. One could expect CK to have **cʔ-* so as to fit in better with the Hadza/Sandawe forms, but if PCK **cʔ-* indeed = **ckx-*, as has been demonstrated in II.78, glottalisation in Hadza and Sandawe does not have anything to do with glottalisation in CK.

The other match here, between PNK and #Hoan (possibly also PKw, although the form itself is only present in |Xam), looks extremely promising due to the near-total similarity between the two - everything coincides except for the click influx. PNK **!̣-* is known to correspond to #Hoan */-* (II.13) and to

#Hoan hushing sibilants (II.17; II.37; II.94); its ties with the labial click were so far established only for SK (II.53), but cf. also #Hoan _Θoa_ "to kill (pl. action)" - PNK _*!ʔóá_^{ŋ} id.; apparently, in some cases the same correspondence also works for #Hoan.

II.85, II.86. "THAT/THIS":

A) PNK __*Tòʔá__ (Žu. _tòʔá_ "that", !Kung (Lloyd) _doa_ "this, that", !O. _doa_ "that"); PT __*tV__ (!Xóõ _tVʔV_ "this", _tàʔa BV kV_ "that", Mas. _ta_, _ti_ "that", _te a_, _ti e_ "this", |Nu//en _ti_ "that"); PKw __*tV__ (//Kxau _ti_ "this", //Ku//e _ti_ "this, that", |Auni _ti_ "that"); #Hoan __ćoa__ "that"; B) PNK __*čV__ (//Au. _či_ "this, that", !O. _či_ "this"); C-D) PNK __*hè__ (Žu. _hè_ "this"); PT __*ʔV__ (!Xóõ _ʔVV_ "this, that"; PKw __*(H)a/*(H)e__ (|Xam, //Ng, |Auni _a_ "this, that", |Xam _ha_, _he_ "this, that", Bat. _ha_ "this, that", |Auni _ha_, _hi_ "that"); #Hoan __ha__ "this"; PNKK __*a__ "that" (//Ani _ʔá-tè_, Kxoe _ʔá_, //Gana _ʔá-séá_, #Haba _ʔà-sá-hà_, Hie. _a_); PNKK __*i__ (Kxoe _ʔí_ "that (fem.)", Naro _ʔì-sí-hà_ "that", Deti _hí_^{ŋ} "this", Cara, Danisi, Kua, Tsua _i_ "this"); Sandawe __ha__ "that", __he__ "this"; Hadza __ha__ "this"; E) PKK __*ne__ "this" (Nama _ne_); PNKK __*ǀné__ "this" (//Ani _ǀné_ "this", _ǀné-tè_ "that", Buga _ǀné_ "this", _ǀne-hà_ "that", Kxoe, Naro, //Gana, #Haba, Tsixa _ǀné_ "this", Kua, Tsua _ǀní_ "this"); F) PKK __*//na__ "that" (Nama _//ná_^{ŋ}, !Ora _//na_); PNKK __*//na__ (Kxoe _//ǀnǎ_ "that", Naro _//ǀnà_ "this").

Notes. It is extremely hard to deal with Khoisan demonstrative pronounts separately - practically none of the subgroups seem to draw a sharp line between the bases for "this" and "that", either switching functions between two bases in what seems a thoroughly random fashion, or using one root for both, with the difference in meaning expressed with a suffix, vowel alternation, or, sometimes, merely a tonal opposition. In view of this, it will be more appropriate to discuss both words together.

Apart from more "local" cases like the NK stem _*čV_ or the CK stems _*ǀne_ and _*//na_, Khoisan shows three distinctly opposed stems: _*TV_ (NK and SK), _*a_ (omnipresent), and _*e/*i_ (also omnipresent). Note, however, that the latter two bases are often undistinguishable from each other - in !Xóõ, for instance, the vowel quality in the demonstrative pronoun is dependent on the phonetic/morphologic characteristics of the adjacent noun. All of these three bases probably date back to Proto-Khoisan and were used to indicate various levels of deixis, but for now, it is hardly possible to assign them exact meanings, considering how little we know of the actual deictic system in the majority of Khoisan languages. External data from other language macrofamilies indicates that the stem _*a_ is normally used for "that" and _*i_ (_*e_) for "this" all around the world, and this is, in fact, the situation that we find in Sandawe; on the other hand, Sandawe seems to lack the _*t_-form so widespread in other branches.

The following conclusions are the most important to our lexicostatistical analysis: a) the _*t_-stem is characteristic of NK, SK, and #Hoan, but _not_ any of the other branches; b) the two vocalic stems, whatever their exact meanings were, are definitely opposed in CK and Sandawe, but have most probably merged everywhere else; c) all the other stems are either innovations in respective branches, or have ceased to function as the main forms for "this" and "that" on the protolanguage level in the branches where they are not found.

II.87. "THOU":

A) PNK __*à__ (//Au. _a-hi_, Žu. _à_, !O. _a-hi_, _a_); PT __*ā__^{h} (!Xóõ _ā_^{h}, Mas., |Nu//en _a_); PKw __*a__ (|Xam, //Ng, #Kho., //Kxau, Bat., |Auni, |Nusan _a_); PKK __*sa-c, *sa-s__ (Nama _sá-ts_ (m.), _sá-s_ (f.), !Ora _sa-c_ (m.), _sa-s_ (f.)); PNKK __*cá__ "thou (m.)" (//Ani, Buga, |Ganda, Naro, Deti, Cara, Tsixa, Danisi, Cua, Kua, Tsua _cá_, Kxoe _tçá_, |Gwi _cí_, //Gana, #Haba _câ_, |Xaise _ćá_, Hie. _ča_); PNKK __*sá__ "thou (f.)" (//Ani, Buga, |Ganda _há(_^{ŋ}_)_, Kxoe _há_^{ŋ}, |Gwi _sí_, //Gana, #Haba _sâ_, |Xaise, Deti, Cara, Tsixa, Danisi, Cua, Kua, Tsua _sá_); B) #Hoan __u__; C) Sandawe

hapu; D) Hadza ***te***.

Notes. PCK is the only language for which two different stems of the 2nd person sg. have to be reconstructed - masculine **cá* and feminine **sá* (Hadza also distinguishes between masculine and feminine forms, but only by means of an additional suffix - *te* m., *te-ko* f.). A closer analysis shows that the same opposition (**-c-* for masculine, **-s-* for feminine) is characteristic for a number of other pronominal forms, cf., for instance, //Ani *com* "we" (excl. masc. dual), *som* "we" (excl. fem. dual), etc.; it turns out that these are actually detachable gender prefixes, easily reconstructible on the PCK level. As for the PKK forms, they are obviously descended from **sa-ca* and **sa-sa*, with reduction of the second syllable; the initial **sa-* is the "full stem indicator" morpheme that gets lost, for instance, when the pronoun is used in the object position.

In this way, PCK shows the same basic "pure" stem **-a-* for the 2nd sg. pronoun as NK and SK. This is an extremely important isogloss and a very strong argument in favour of Khoisan relationship, as no other major language family in the world is known to have **a* in this particular function. Within "South African Khoisan" only #Hoan shows a different stem - possibly influenced by the 2nd plural stem **u*, found in SK. Whether the same **-a-* is represented in Sandawe *h-a-pu* is unclear; however, considering that the indirect (enclitic) stem in Sandawe is simply *-pV-*, and that *ha-* also functions as a potential prefix in other pronominal forms (*ha-we* "he", *ha-su* "she"), we cannot propose a match here. As for Hadza, it shows a definitely different stem, closer to Euroasiatic, in fact, than to any other macrofamily.

II.88. "TONGUE":

A) PNK ****dharì*** (//Au. *tari*, Žu. *dharì*, !O. *tali*); PT ****?/nạ̀*** (!Xóõ *?/nạ̀n*, Mas. */nan*, |Nu//en */ani*); PKw ****/?a-***, ****/?e-***, ****/?a-Ri*** (|Xam */érri*, */énni*, //Ng */e^{p}*, //Kxau */?anansi*, |Auni */a^{p}ri*); #Hoan ***cela***; PKK ****dlamV*** (Nama *nammi*, *lammi*, *tammi*, !Ora *támma*); PNK ****dàḿ*** (//Ani, Buga, Naro, |Xaise, Deti, Cara, Tsixa, Danisi, Cua, Kua, Tsua *dàḿ*, |Ganda *dàm̀*, |Gwi *gjàm̀*, //Gana, #Haba *dáḿ*, Hie. *dham*); Hadza ***/natha-***; B) Sandawe ***!he^{p}***.

Notes. The root "tongue" is known to behave irregularly in quite a few language families over the world, and it is hardly a pure coincidence that the irregular fluctuation of consonantism in Nama (*nammi*, *tammi*, *lammi*) is strangely similar to the one observed in, for instance, Latin (*dingua*, *lingua*); in both cases we may be witnessing either a unique consonant cluster in the protolanguage, or a specific irregularity characteristic of this particular word. Within Khoisan, the same peculiar behaviour can be observed for the root "throat" (PNK **do^{h}m*; PT **?/nụ́m*; PKw **dọm*; #Hoan *?/nọo*; PCK **dóm̀*), although the reflexation of the original phoneme/cluster is slightly different. The lexeme is represented in NK, SK, and (possibly) #Hoan with a **-rV/*-nV/*-lV* suffix, as opposed to CK **-m*, but the root itself is most probably the same everywhere - only #Hoan *cela* is dubious, considering that for "throat" #Hoan has *?/n*, just like !Xóõ. Hadza */natha-* can also tentatively belong here due to a direct phonetic match with PT; Sandawe is the only language that clearly has a different root, as there is no way to link an initial *!h-* to all these other reflexations.

II.89. "TOOTH":

A) PNK ****c?àù*** (//Au. *cau-si*, Žu. *c?àù*, !O. *cau*); #Hoan ***ciu***; B) PT ****//qhV̀^{p}*** (!Xóõ *//qhà^{p}*, Mas. *//xu^{p}*, |Nu//en *//Vn-te* pl.); PKw ****//khe^{p}*** (|Xam *//khe^{p}*, //Ng *//a^{p}*, *//e^{p}*, #Kho., Bat. *//e^{p}*); PKK ****//u^{p}*** (Nama *//gû-b*, !Ora *//ú^{p}-b*); PNKK ****//ú^{p}*** (Kxoe, |Gwi, //Gana, |Xaise, Cara, Tsixa, Danisi, Cua, Tsua *//ú^{p}*, Buga, Naro, Deti *//ú^{p}*, #Haba *//ụ^{p}*, Kua *//û^{p}*, Hie *//goo*); C) Sandawe ***!?akha^{p}***; D) Hadza ***?áhà-***.

Notes. Another obvious match between NK and #Hoan here. The connection between SK and CK is somewhat less obvious because the vocalism correspondences are blurred, but on the other hand, the correspondence "PT **-q(h)-* - PCK zero" is practically the same as in "claw" (II.13), which leaves no unresolved problems in the consonant area.

II.90. "TREE":

A) PNK ***!ài**^{hŋ}**~*!àu**^{hŋ} (//Au. *!gai*^{ŋ}, *!ga*^{ŋ}, Žu. *!ai*^{hŋ}, !O. *!gau*^{ŋ}, *!ga*^{ŋ}); B) PT ***ʔΘnà** (!Xóõ *ʔΘnàje*, Mas. *Θnoe*, |Nu//en *Θʔa*); PKw ***Θho** (|Xam, Bat. *Θho*, //Ng *Θgo*, *Θho*, #Kho. *Θgo*, //Kxau *Θo*, |Auni *Θgoa*, *Θo*); C) #Hoan ***/ʔσ**^{ŋ}; D) PKK ***hai** (Nama *hai-i*, !Ora *hai-sa* "bush"); PNKK ***ji** (//Ani, Kxoe, Buga, Naro, #Haba *jǐ*, |Ganda, |Gwi, //Gana, Deti, Cara, Tsixa, Danisi *jì*, |Xaise *ʒî*, Kua *jî*, Hie. *hii*); E) Sandawe ***thē***, E) Hadza ***çiti-***.

Notes. No matches between any of the major sub-branches, although the question of #Hoan */ʔσ*^{ŋ} vs. the SK forms with the labial click needs further investigation - so far, it is only clear that #Hoan and SK labial clicks mostly stem from different origins and are not directly related to each other, but there is too little #Hoan data to uncover the real picture. Since #Hoan mostly agrees with NK in terms of click correspondences, it would be reasonable to assume that #Hoan */* can correspond to PSK **Θ* if the same correspondence can be found between PNK and PSK; so far, however, the more frequent correspondence is PNK **!̯* to PSK **Θ*.

II.91. "TWO":

A) PNK ***cá̯**^{ŋ} (//Au., !O. *ca*, *ča*, Žu. *cá̯*^{ŋ}); B) PT ***#nûm** (!Xóõ *#nûm*, Mas. *//num*, *!num*, |Nu//en *!num*); PKw ***!ʔu~*//ʔu** (|Xam, //Ng *!u*, *!ʔu*, #Kho., //Kxau, //Ku//e *!ʔu*, Bat. *//u*, *//ʔu*, |Auni *!ʔu*); C) #Hoan ***Θoa***; D) PKK ***/am** (Nama */gam*, !Ora */am*); PNKK ***/ám̀** (all languages have */ám̀*); E) Sandawe ***ki***, ***ki-soxo***.

Notes. No obvious matches here. It is curious, however, that this is the second time in the wordlist where PKw yields a peculiar alternation of alveolar and lateral clicks as a correspondence to PT **#*; the first time was also a numeral ("one", II.63), and in both cases the respective form in #Hoan begins with a labial click. Normally, we should expect PNK and PSK **/* where #Hoan has **Θ* (cf. II.38), but since we do not exclude the possibility of more than one-to-one correspondences, this case is very much in need of further investigation, although it requires more #Hoan data, as well as a more detailed analysis of the *!~//* alternation in the !Wi subgroup.

II.92. "WALK (GO)":

A) PNK ***ú** (//Au., !O. *u*, Žu. *ú*); PT ***û** (!Xóõ *û̯-lu* "to enter, go into", Mas. *u*); PKw ***u** (|Xam, //Kxau *u*, //Ng *u* "to pass", Bat. *u* "to go away"); B) #Hoan ***cáo***; C) PKK ***!u**^{ŋ} (Nama *!gû*, !Ora *!u*^{ŋ}); PNKK ***!û**^{ŋ} (//Ani, Kxoe, Buga, |Ganda, //Gana, #Haba, |Xaise, Deti, Cara, Tsixa, Danisi, Kua, Tsua *kû*^{ŋ}, Naro, |Gwi *!û*^{ŋ}, Hie. *kho*); D) Sandawe ***hiḳi***, Hadza ***haka-***.

Notes. The root **u* "to walk, go, go away" is widely represented in both NK and SK, but does not seem to be present anywhere else. On the other hand, this is compensated by an interesting isogloss between Hadza and Sandawe - provided we can disregard "extra" glottalisation in Sandawe.

II.93. "WARM":

A) PNK ***//ʔú**^{ŋ} (Žu. *//ʔú*^{ŋ}); #Hoan ***//qʔσ***; PNKK ***//ʔô** (Kxoe *//ʔȯ*, Buga, Naro, //Gana, #Haba *//ʔô*); B) PT ***kûbi** (!Xóõ *kûbi*); PNKK ***khóbò** (//Ani, Buga, |Ganda *khóbò*, Kxoe *kxóbò*, Naro *khóbò* "to sweat"); C) PKK ***!nae*** (Nama *!nae-sa*); D) Sandawe ***hoḳa***.

Notes. The first root (**//q?U*) is a good match between NK, #Hoan, and NKK, with the ejective uvular efflux lost in those branches where it is never actually found, predictably replaced by a glottal stop. There is also a more local click-less root, represented by !Xóõ *kûbi* and West Central Khoe **khóbò*; despite its limited distribution, the possibility of borrowing is rather low due to minor phonetic disagreements between branches.

II.94. "WATER":

A) PNK ***!̨gú*** (//Au. *//gu*, *!gu*, Žu. *!gú*, !Kung (Doke) *!̨gu*^{ɒ}, !O. *//gu*); PT ***!qhà*** (!Xóõ *!qhàa*, Mas. *!kha*, *!xa*, |Nu//en *!kha*); PKw ***!kha*** (|Xam *!kwa*, *!khwa*, //Ng *!ha*, *!kha*, #Kho. *!kha*, Bat. *//kha*, |Auni *//kha*, *!a*); #Hoan ***žo***; PNKK ***chǎ*** (//Ani, Buga, |Xaise, Deti, Cara, Tsixa, Danisi, Kua, Tsua *chǎ*, Kxoe *çǎ*, |Ganda, Naro *chà*, |Gwi, //Gana *châ*, #Haba *chá*, Hie. *caa*); Sandawe ***ça***; B) PKK ***//ami*** (Nama *//gammi*, !Ora *//ammi*); D) Hadza ***?athi-***.

Notes. This root, once again, returns us to the problem of the correlation between #Hoan hushing sibilants, NK retroflex clicks, and CK affricates. It is very probably related to "rain" (II.65) - in SK the same form often has both meanings - and features correspondences similar to or coinciding with the ones established earlier in II.17, II.37, and II.65, with the addition of Sandawe *ç-* - provided the similarity between Sandawe *ça* and PNKK **chǎ* is not coincidental; however, given the several satisfactory cases where CK has *ch-* corresponding to NK retroflex clicks, I actually feel even more inclined to group CK "water" with NK and #Hoan "water" rather than with the respective Sandawe item. That said, further data is necessary to confirm this particular grouping, particularly data on #Hoan *ž-* and PNKK **c(h)-*.

II.95. "WE":

A) PNK ***è*** excl. (//Au. *e*, Žu. *è*, *è-!á*, *è-cá* dual, !O. *e*, *ehŋ*); PT ***í***^{h} (!Xóõ *í*^{h}, Mas. *i*, |Nu//en *i*, *e*); PKw ***i*** incl. (|Xam, //Ng, //Ku//e, Bat., |Auni *i*); PKK ***-e*** plural (Nama *sí-k-e* m. excl., etc.; !Ora *si-tj-e* m. excl., etc.); PNKK ***-e*** plural; B) PNK ***m*** incl. (Žu. *m*, *m-!á*, *m-cá* dual, !O. *m*, *m-!a*); #Hoan ***ǹ-!à?e***, PKK ***-m*** dual (Nama *sí-kx-m̀* m. excl., etc.; !Ora *si-kha-m* m. excl., etc.); PNKK ***-m*** dual; C) PKw ***si*** excl. (|Xam, //Ng, //Kxau, //Ku//e, |Nusan *si*, #Kho. *si*, *sa*, |Auni *si*, *se*); D) Sandawe ***su***^{ɒ}; Hadza ***u-***.

Notes. Most Khoisan subgroups distinguish between exclusive and inclusive 1st person plural/dual; both stems are included into comparison where present. The actual pronominal system can be extremely complex in this case, particularly in CK, where nearly every form of the 1st person plural/dual pronoun consists of three distinct (or not so distinct) morphemes, one indicating the "full form" of the stem (which disappears in the object position), one marking the class/gender, and one actually serving as the main pronominal base; thus, in Nama *sí-kx-m̀ si* is the "full stem marker", *-kx-* the masculine gender marker, and *-m̀* the main pronominal stem for the dual number. This system, naturally, undergoes a lot of changes in different CK languages, and it is hardly possible to go into the details of these changes because of space limitations; a complete list of forms and paradigms, as well as ample historical commentary, can be found in [Vossen 1997].

The most important detail of the CK system in general, however, is that one can firmly establish two main pronominal morphemes, **-m* for 1st p. dual and **-e* for 1st p. plural. These bear a striking resemblance to, respectively, PNK **m* "1st p. inclusive" and PNK **è* "1st p. exclusive"; apparently, either the opposition "dual/plural" had been at one time reinterpreted in PNK as "inclusive/exclusive", or vice versa. A further clue here can be found in comparing PKw **si* "1st p. inclusive" with the PKK "full stem marker" **si-*, used *only* in 1st p. inclusive pronouns; this can indicate that originally it was the morpheme **si-* that was used to express the idea of "inclusiveness", while **-m* and **-e* were primarily differentiated by the category of grammatical number. This is why I am not counting Nama *sí-kx-m̀* and PKw **si* as an exact match - because if the morpheme **si-* really goes back to Proto-Khoisan, it functions rather as a special pronominal prefix than a true pronominal root.

As for the rest of the forms (that is, the ones that are not **m*-type or **-e-*/**-i-*type), #Hoan *ń-!à'e* is probably related to PNK **m*, with some sort of contextual sandhi before the pronominal suffix (which is common to both #Hoan and PNK), and both Hadza and Sandawe agree in having an *u-* stem that is absent in other Khoisan languages.

II.96, II.98. "WHO/WHAT":

A) PNK ***(h)à*** (Žu. *ha-žòè* "who", *ha-čé* "what"); PKK ***ha-*** (Nama *hám* "who", !Ora *ham* "which"); PT ***é***^{h} (!Xóõ */V... éh* "who, what"); B) PKw ***d(j)e*** "who, what" (|Xam *de*, //Ng *gi-si*, *ki-si*, #Kho. *gji-si*, //Kxau *de*); #Hoan ***ćini*** "what"; PKK ***TV*** (Nama *tari-ʔi* "who", *tare-ʔi* "what", !Ora *da-* "who"); PNKK ***(n)dV*** (Naro *di* "who", *du* "what", |Gwi *dí* "who", //Gana *dí* "who", *dǔ* "what", Deti *dú* "what", Cara, Danisi *ndú* "what"); Hadza ***tama*** "who, what", ***da*** "who"; C) PNKK ***ma(***^{ŋ}***)*** (//Ani, Buga *ma* "who", Kxoe *má*^{ŋ} "who", Naro *m* "what", |Gwi *mâ*^{ŋ} "what", //Gana *mâ* "what", Deti, Cara, Kua *má* "who", Tsixa, Danisi *má(é)* "who"); D) PNKK ***nV*** (//Ani, Tsixa *né* "what", Buga *ná* "which", Cara *ná* "who", Kua *nǎ* "who", *ná* "what"); E) Sandawe ***ho*** "who", ***ho-co*** "what".

Notes. Judging by the material, any opposition between the stems for "who" and "what" found in Khoisan should be judged secondary; the majority of the branches simply does not have two different stems, expressing the difference by means of affixation, tonal distinction (as in Kua), or not expressing it at all (as in !Xóõ). Thus, the two items of the wordlist have to be discussed together, just as in the case of Khoisan demonstrative pronouns.

NK and SK agree in showing a monovocalic interrogative base (**a~*e*, the original vowel is impossible to determine, especially since the original vocalism could have been obscured by later assimilation processes caused by the morpheme's clitical status), which, within CK, is only present in PKK. The other important morpheme is **TV*, widely spread in !Wi and in all CK branches, but practically absent in NK and Taa languages (at least, according to recorded data). Apart from that, CK also boasts two additional interrogative morphemes, **ma* and **nV*, none of which have any direct correlates outside CK. Hadza and #Hoan seem to agree with !Wi and CK in having **TV* as the main stem (#Hoan *ćini* < **tV-ni*); as for Sandawe *ho*, it remains yet to be seen if we can actually compare it to the NK/SK monovocalic stems, since there is no reason for a 'secondary' labialisation.

II.97. "WHITE":

A) PNK ***!àʔú*** (//Au., !O. *!au*, Žu. *!àʔú*); PKw ***!ʔu(i)*** (|Xam *!ui-ta*, #Kho. *!ʔu-rija*, //Kxau *!ʔui*); PKK ***!ʔu-ri*** (Nama *!uri*); PNKK ***!ʔú*** (Naro, |Gwi, //Gana, #Haba, Tsua *!ʔú*, Kua *!ʔǔ*); B) PT ***!nú*** (!Xóõ *!nú-ná*); C) #Hoan ***#ạʔana***; D) Sandawe ***phō***; E) Hadza ***peƛ̣a-***.

Notes. A good match between PNK, PKw, and both CK branches; the glottal stop efflux in CK is almost certainly related to the inlaut glottal stop in NK, which explains the formal lack of direct correspondence between effluxes, although it is unclear which variant is primary here. Sandawe and Hadza cannot be counted as matches, since there is no further evidence for Hadza *-ƛa-* being a suffix of any sort.

II.99. "WOMAN":

A) PNK **_*ʒhʔàú_** (//Au. _ʒau, ʒou_, Žu. _ʒhʔàú_, !O. _cau_); B) PKw **_*/ʔai-tje_** (|Xam _/ʔai-tje, /ʔai-ti_, //Ng _/ʔai-ti, /ʔai-ki_, #Kho. _/ʔai-tje_, //Kxau, //Ku//e _/a-ti_, Bat. _/a-ze_, |Auni _/ge-kı_); C) #Hoan **_//gai, ʔari-//gai_**, PNKK **_*//gae, *//gae-khoe_** (//Ani _//gɛ-_, Kxoe _//gɛ̌-_, Buga, |Ganda, //Gana _//gɛ-_, Naro, |Gwi, #Haba, |Xaise, Deti, Danisi, Kua, Tsua, Hie. _//gae-_, Cara, Tsixa _//ga-_); D) PKK **_*tara_** (Nama, !Ora _tara-s_); Sandawe **_thame-cu_**; E) Hadza **_ʔakhwiti-_**.

Notes. While some of the groups have a distinct root for "woman" (NK, !Wi), others either replace it with the root "female" (#Hoan _//gaı_), or use the compound "female person", like PNKK. In the Taa subgroup the exact proto-form cannot be established, since some languages use the form _*ta-qae_ (lit. "person-mother"), while others prefer _*ta-//gae_ ("person-female"). #Hoan displays a rare case of agreeing with CK (rather than NK) in having _//gae_ as the main root for "woman". There is also an interesting possible isogloss between PKK and Sandawe (both _-ra_ and _-me_ can be seen as fossilized suffixes, as they're frequently found in other nouns).

II.100. "YELLOW":

Almost none of the sub-branches of Khoisan allow for a distinct root with the meaning "yellow" to be reconstructed. Most languages either use the same word for "yellow" and "green", or, on the contrary, distinguish between several different shades of yellow, making it impossible to determine the "main" root; finally, in some groups/languages the word "yellow" has not even been recorded. We will have to exempt this particular item from the analysis.

III. **CONCLUSIONS.**

III.1. _Phonetic correspondences._ It is practically obvious from this analysis that any attempt to reconstruct Proto-Khoisan based on a "one-to-one" correspondence system is bound to fail. Out of all the items in the wordlist, only a miserably small minority of cases demonstrates such correspondences (that is, when _both_ click influxes and effluxes are taken into account) - and some of these cases, such as "fish," are highly suspicious in the first place. On the other hand, while some of the proposed "non-trivial" correspondences are supported by more data than others, it seems clear to me that the true picture can be established only by attempting to go beyond the concept of "obvious resemblance."

Let us list all the suggested click influx correspondences between the "main" branches of Khoisan, excluding Hadza and Sandawe (since PNKK and PKK data never disagree with each other about the click influx treatment, both will be joined here under PCK; same with PT and PKw, joined under PSK):

(a) PNK */- PSK */- #Hoan /- PCK */(II.9, II.20, II.28, II.33, II.35, II.46, II.62, II.82);
(b) PNK */- PSK *#- #Hoan /- PCK *#(II.21, II.74);
(c) PNK */- PSK */- #Hoan _Θ_ (II.25, II.38, II.63);
(d) PNK *#- PSK *#- #Hoan #- PCK *#(II.2, II.8, II.18, II.45);
(e) PNK *#- PSK *//- #Hoan #(II.15, II.80);
(f) PNK *!- PSK *!- #Hoan !- PCK *!(II.4, II.34, II.41, II.43, II.97);
(g) PNK *!- PSK */- #Hoan !- PCK */(II.36, II.40, II.57, II.66, II.81);
(h) PNK *!- PSK *//- PCK *//(II.11, II.44, II.79);
(i) PNK *!̣- PCK *!(II.45);
(j) PNK *!̣- PSK *//- #Hoan !- PCK *//(II.13, II.54);

(k) PNK *!̭ - - PSK *! - #Hoan š~ž - PCK *ch (II.17, II.33, II.37, II.65, II.94);
(l) PNK *!̭ - PSK *Θ - (?) #Hoan Θ (II.53, II.84);
(m) PNK *// - PSK *// - #Hoan // - PCK *// (II.13, II.29, II.41, II.89, II.93, II.99).

Some of the "one-to-one" correspondences seem to be supported by a significantly bigger amount of data than others, but a closer look reveals that these are mostly cases of PNK vs. #Hoan correspondences, and these two branches never disagree with each other when it comes to click influxes, with the exception of the #Hoan labial click. Apart from that, practically every one of these correspondences finds at least some support from outside data, and the strict semantic criteria applied to data selection significantly diminish (though by no means liquidate) the probability of chance resemblances.

Even from this limited amount of data we can already draw some conclusions. First, as has already been mentioned, is the peculiar "agreement" between PNK and #Hoan. These two branches, on the other hand, are obviously opposed by PSK and PCK, which only disagree when it comes to SK labial clicks and the "fricatives vs. clicks" correspondence (k). Which situation should be considered original - the PNK/#Hoan one or the PSK/PCK one - is, of course, an open question. Either the original PK system was preserved in PNK/#Hoan, with certain unknown factors contributing to the bi- and trifurcation of click reflexes in the other two branches, or, vice versa, the original system underwent radical changes in PNK/#Hoan, with factors that "forced" most of the clicks to merge in the alveolar one in a large part of the lexicon.

It must be specially noted that the correspondences given above in no way violate the Neogrammarian demand of *regularity* of phonetic changes, for one simple reason: so far, I am not proposing any actual *reconstructions* of Proto-Khoisan phonemes that would yield an unexplained bifurcation of reflexes in daughter languages. All of these bifurcations have to have some kind of explanation - but it must be remembered that there might have been additional factors at work in Proto-Khoisan itself that have not survived in any of the daughter languages. There might have been tonal, prosodic, vocalic, or other reasons that have progressively become obscured over time; there might also have been additional factors of click articulation (for instance, "tense" and "lax" articulation, or labialized/non-labialized articulation) that have become neutralized in PNK, PSK, and PCK after the original split. In any case, a set of major phonological changes in sub-branches of Khoisan, taking place over a period of seven or eight millennia, seems to me a much more realistic probability than a set of minor "cosmetic" phonological changes in these sub-branches over an even longer period (which would have to be assumed based on glottochronological calculations according to the "one-to-one" correspondence principle).

It could thus be argued, in terms of historical typology, that the situation with Khoisan historical phonetics might well be similar to that of, for instance, the Proto-West-Caucasian system. In the latter case, while the actual modern day phonological systems of West Caucasian languages such as Abkhaz and Adygh, already quite rich and complex by themselves, are quite close to each other, the correspondences between them are of an extremely complex nature and betray a proto-system even richer and more complex in oppositions than any of its daughter languages. (Cf., for instance, six distinct series of affricates based on the three oppositions of "hissing/hushing," "palatalized/non-palatalized," "labialized/non-labialized," not more than four of which survive in daughter languages.) This kind of interpretation currently seems to me to be the most promising one, and should indicate the direction of further studies.

We must also keep in mind another extremely important factor: the obvious "imperfection" of our knowledge of Khoisan phonetics. So far, out of the three major subgroups of Khoisan only Central Khoisan languages have received a generally satisfactory description, and it is generally acknowledged among those who share the belief in Common Khoisan that Central Khoisan actually seems to simplify the original system to a much higher degree than the other branches. For North and South Khoisan, on the other hand, we only have exhaustive descriptions for *one* language/dialect per branch, and it is quite possible that, despite the complexity of their phonetic systems, there could be significant phonological oppositions that eventually became neutralized in these very languages (Žu|'hoan and !Xóõ).

In fact, at least for Žu|'hoan this is exactly the case. As has been demonstrated above, this dialect has lost the important phonological opposition between the alveolar and the retroflex click. As for South Khoisan, let us not forget that the dictionary of Anthony Traill has more than *doubled* the number of click effluxes that had been identified by previous explorers of South Khoisan, and has also introduced click effluxes that nobody has *ever* identified before in *any* Khoisan language.

Even for Central Khoisan the situation is far less clear than may be deemed at first. As has been shown in [Vossen 1992], for instance, Central Khoisan is not entirely devoid of uvular click effluxes, and it remains to be seen how well they correlate with uvular effluxes in other Khoisan languages. Vossen also postulates a newly found phonological opposition between the "voiced nasal click" (*|ᵑn*, etc.) and the "voiceless" nasal click (*/n*, etc.) in many Non-Khoekhoe languages; this opposition has not been taken into account in the present work yet, but it certainly deserves further attention. And in a relatively recent paper by Hirosi Nakagawa [Nakagawa 1996], yet another previously unheard of opposition is suggested for the language of |Gwi - between a simple glottal stop efflux (*/ʔ*, etc.) and a "voiceless velar ejective" efflux (*k/ʔ*, etc.), which the author claims is not only discernible acoustically and found in minimal pairs, but is even confirmed through external comparison with the language's closest relative, //Gana.

All of this goes to show that the "untrivial" correspondences, found by us among click effluxes as well as click influxes, may, in fact, not only reflect oppositions that have been lost in modern languages, but may actually reflect oppositions that have not yet been found. This is why further research and field work on poorly described Khoisan languages is of such crucial importance for any attempts to arrive at a reconstruction of Proto-Khoisan.

III.2. *Classification.* The other important result of our lexicostatistical analysis is that it allows shedding more light on the problem of the relation between various branches of Khoisan. Calculation of the percentage of matches between these branches helps us affirm that at least some of these branches are, indeed, related, and that the Khoisan genealogical tree should be treated as a rather complex hierarchic structure. Below I am listing some of the calculation results, going from higher numbers to lower ones to illustrate the various levels of relationship.

(a) Level 1 (65-70% matches). This is the relationship between PKw and PT (68%); and PNKK and PKK (70%). These actual relationships have never really been put in much doubt, and lexicostatistics/glottochronology only further confirms what has been long before stated given the number of phonological, lexical, and grammatical isoglosses between these branches. The huge number of matches suggests that the actual split between the two main branches of South Khoisan, on one side, and Central Khoisan, on the other side, took place sometime in the 1st millennium B.C.

(b) Level 2 (40-45% matches). This is the case of PNK vs. #Hoan (43%). The resemblance between #Hoan, on one hand, and different NK and SK languages, on the other hand, have been noticed long before (see, for instance, [Traill 1973], where this resemblance is reflected in the very name of the article); since then, some researchers, chief among them Henry Honken, have preferred to link #Hoan more closely to NK, even including it directly in NK. As the above percentage shows, lexicostatistical analysis more or less confirms this hypothesis. #Hoan can be safely assumed to represent an "elder brother" of modern NK dialects, much more distant from them than they are from each other, but significantly closer to NK than anything else. According to glottochronological calculations, the split of "Proto-NK-#Hoan" must have taken place somewhere around the 2nd millennium B.C.

(c) Level 3 (30-35% matches). PNK has 32% matches with PT and 34% matches with PKw; #Hoan shows 28% matches with PT and 21% matches with PKw (the latter figure could be seriously erroneous due to the fact that PKw and #Hoan wordlists are the most incomplete and questionable of all). This confirms that NK and SK subgroups are actually more closely related to each other than to CK, an idea that is also supported by the fact that the grammatical systems of both groups share more similarities than any of them with CK, including more similar pronominal systems and a lack of complex verbal morphology, among other features. Glottochronology shows that the split between NK/#Hoan and SK must have taken place somewhere around the 4th millennium B.C.

(d) Level 4 (20-25% matches). PNK has 19% matches with PKK and 26% matches with PNKK; PT - 15% with PKK, 23% with PNKK; PKw - 20% with PKK, 24% with PNKK. These figures might be slightly lower if we eliminated several possible loans from PCK into other branches, like "fish," etc. Nevertheless, the figures still suggest that what has often been called "Proto-South-African Khoisan," i.e. the ancestor of modern day NK, SK, and CK languages, underwent its first split - into Proto-NK/SK and Proto-CK - some time around the 6th millennium B.C., i.e. is much older than, say, Proto-Indo-European, and should rather be compared in depth to deeper families such as Altaic.

(e) Level 5 (8-12% matches). This is the case of Hadza and Sandawe. The actual number varies from as much as 14% (Sandawe and PT) to as little as 2% (Sandawe and #Hoan). However, this is obviously the weakest spot in all the calculations. Most of the ties between Hadza/Sandawe and other languages have been established based on *resemblance* rather than *correspondence*, and, although a few parallels (like Sandawe _/i_ - PNK _*ci_ "come", Sandawe _ƛana_ - PCK _*//na_ "horn", etc.) look quite promising, even these are only vaguely confirmed by additional data. One thing is for certain: even if the figures are *approximately* true - which would put the split of "Proto-Khoisan" or "Proto-Macro-Khoisan" somewhere around the 8th or 9th millennium B.C. - and Sandawe and Hadza are indeed closer to the rest of Khoisan than to any other language family, our chances of arriving at a plausible reconstruction of this "Proto-Macro-Khoisan" look rather grim, comparable with, for instance, trying to reconstruct Proto-Nostratic on the basis of Proto-Altaic, modern day English, and modern day Hungarian.

That said, the final answer to this question must be postponed until at least two things have been done: a) a careful lexicostatistical analysis of Hadza and Sandawe data vs. at least some of the *other* African language families, so as to be sure that their ties with Khoisan are indeed closer than with anything else, and b) a working reconstruction of "Proto-South-African Khoisan". If anything, the results of these calculations, as well as the established correspondences, make me hope that such a reconstruction is, indeed, possible, and, given time and additional language material, will eventually appear.

NOTES ON TRANSCRIPTION

There is no universally accepted system of transcribing the extremely complex phonological inventories of Khoisan languages; apart from a general agreement about the symbols used to denote most click influxes, each researcher usually has his or her own preferences (and since new phonological oppositions are being discovered or at least suggested regularly, these preferences also change from time to time). Keeping in mind that the present work might be of interest to specialists outside the general field of Khoisan linguistics, I have tried to unify all the numerous transcriptions as well as possible in order to relieve the reader of the necessity of consulting the original sources for explanations.

It should be noted that the unification of transcription is far from an easy task, and in some cases is almost impossible to do, especially when dealing with older sources suffering from particularly imprecise phonological transcription. For example, Dornan records the Hietšware form for the verb "kill" as */goo*; obviously, it is the same root with the general CK root for this verb, normally marked by Vossen as */kxúᵑ*; according to the laws of CK historical phonetics, the intermediate branch from which the Hietšware form is descended has it represented as */ʔúᵑ*. The question, then, is whether the Hietšware form *indeed* had a voiced efflux *-g-* and denasalisation of the vowel, or if this was merely an error on Dornan's part. Since, upon further analysis, we note that where Kua and Tsua (Hietšware's closest relatives) have *-ʔ-*, Hietšware has either zero, *-k-*, *-g-*, or *-h-*, we may draw the conclusion that Dornan's transcription of click effluxes was essentially misguided, and they should not be taken into account. But does this give us a right to "rectify" his transcription? Unfortunately, no, because knowing that Dornan could be wrong does not shed any light onto what was actually *right*. It could well be that in the dialect he was transcribing, the glottal stop efflux regularly developed into, say, an aspirated efflux, or merged with the voiced efflux, and we have no way of determining it.

Therefore, any attempts to "unify" the transcription of older sources (mainly represented in [Bleek 1956]) should only touch upon the purely *graphic* aspects of the systems used. One such important discrepancy between the transcription of Bleek and her sources, on one hand, and most newer sources, on the other, is that Bleek transcribes the so-called "zero efflux" as *-k-* and the so-called "glottal stop efflux" as zero, while nowadays the zero efflux is usually transcribed as zero and the "glottal stop efflux" as *-ʔ-* or *-'-*. Thus, Bleek's NK */kam* "sun" = my */am*, and Bleek's SK */a* "to die" = my */ʔa*.

Below I list all the most frequent transcription signs that require explanation, with additional comments where needed.

Click influxes. **ʘ** - labial click; **|** - dental click; **#** - palatal click; **!** - alveolar click; **//** - lateral click; **!̥** - "retroflex" click (in var. sources transcribed as !! or]).

Click effluxes. Zero - "zero" efflux (phonetically with a slightly articulated *-k-*); **ʔ** - glottal stop efflux; **h** - aspirated efflux (in NK phonetically a combination of glottal stop plus aspiration, sometimes transcribed as **ʔh** or **'h**); **kh** - zero efflux with aspiration (only known to be phonologically opposed to **-h-** in NK, sometimes transcribed as **h**); **g** - voiced efflux; **n** - nasal efflux; **ŋ̥** - aspirated nasal efflux (in NK) or voiceless nasal efflux (in SK); **ʔn** - preglottalized nasal efflux. The rest of the effluxes (**x**, **γ**, **q**, **kx**, etc.) represent combinations of clicks with various velar and uvular consonants.

Affricates. The maximum number of affricate series is Khoisan is three, and it is only found in #Hoan, where the original Khoisan dentals developed into palatal affricates. This triple distinction is regularly marked according to the following principle: **c** - hissing series, **č** - hushing series, **ć** - palatal series. Their correlates for other manners of articulation are, for all three series: voiced affricates (**ʒ**, **ǯ**, **ʒ́**), aspirated affricates (**ch**, **čh**, **ćh**), aspirated voiced affricates (**ʒh**, **ǯh**, **ʒ́h**); for two series only - preglottalized affricates (**cʔ**, **čʔ**), preglottalized voiced affricates (**ʒʔ**, **ǯʔ**), fricatives (**s**, **š**), voiced fricatives (**z**, **ž**).

Other consonants. Most of the other symbols used to transcribe consonants are self-explanatory. The uvular series is transcribed in the following way: **q** = voiceless stop, **G** = voiced stop, **q̇** = ejective stop, **qh**, **Gh** = aspirated voiceless/voiced stops. Lateral consonants in Hadza and Sandawe are marked as **λ** (lateral fricative), ƛ (lateral voiceless stop), Ł (lateral voiced stop), and ƛ̣ (lateral ejective stop).

Vowels. In most cases, the transcription is self-explanatory. Occasionally, especially in transcriptions of !Wi data, I use the signs **ɛ** (open *e*) and **ɔ** (open *o*) in the way they are used in the original sources, not being sure of their phonological status in those languages. Pharyngealized vowels are marked as **a̰**, **o̰**, etc. Aspirated vowels ("breathy voice") are marked as $\mathbf{a}^{h}$, $\mathbf{o}^{h}$, etc. Finally, nasalisation is marked as $\mathbf{a}^{ŋ}$, $\mathbf{o}^{ŋ}$, mainly for readability reasons, since otherwise the nasalisation marks could merge with tone marks.

Note on Nama. Nama words are the only ones where the transcription has not been unified, as Nama is currently the only Khoisan language for which there exists an established orthographic norm. It should therefore be kept in my mind that what is marked in Nama orthography as the zero efflux is actually the glottal stop efflux (i.e. *#an* = **#ʔan*); and that the circumflex sign is used in Nama to mark vowel nasalisation, not tone (i.e. *â* = $a^{ŋ}$, etc.).

LIST OF ABBREVIATIONS
Hie. - Hietšware
Mas. - Masarwa
(P)CK - (Proto)-Central-Khoisan
PK - Proto-Khoisan
(P)KK - (Proto)-Khoekhoe
PKw - Proto-!Wi
(P)NK - (Proto)-North-Khoisan
(P)NKK - (Proto)-Non-Khoekhoe
(P)SK - (Proto)-South-Khoisan
PT - Proto-Taa
Žu. - Žu|'hoan
!O. - !O!Kung
#Kho. - #Khomani
//Au. - //Au//en

BIBLIOGRAPHY

Barnard 1985 = *Alan Barnard.* A Nharo wordlist, with notes on grammar. Occasional publications from the Department of African Studies, no 2. Durban: University of Natal. Pp x, 238, 1985.

Baucom 1974 = *Kenneth Baucom.* Proto-Central-Khoisan. In: Proceedings of the 3rd annual conference on African linguistics, 7-8 April 1972, pp 3-37. Ed(s): Erhard F[riedrich] K[arl] Voeltz. Indiana University publications, African series, vol 7. Bloomington: Research Institute for Inner Asian Studies, Indiana University, 1974.

Bell-Collins 2001 = *Arthur Bell, Chris Collins.* #Hoan and the typology of click accompaniments in Khoisan. In: Cornell working papers in linguistics, 18, pp 126-253, 2001.

Bleek 1929 = *Dorothea F. Bleek.* Comparative vocabularies of Bushman languages. Publications from the School of African Life and Language. Cambridge: University Press. Pp 94, 1929.

Bleek 1956 = *Dorothea F. Bleek.* A Bushman Dictionary. American Oriental Series, vol. 41. New Haven, Connecticut, 1956.

Collins 2001 = *Chris Collins.* Aspects of plurality in #Hoan. In: Language, 77 (3), pp 456-476, 2001.

Collins 2001a = *Chris Collins.* Multiple verb movement in #Hoan. In: Cornell working papers in linguistics, 18, pp 75-104, 2001.

Collins 2001b = *Chris Collins.* VP internal structure in Ju|'hoan and #Hoan. In: Cornell working papers in linguistics, 18, pp 1-27, 2001.

Dempwolff 1916 = *Otto Dempwolff.* Die Sandawe: linguistisches und ethnographisches Material aus Deutsch-Ostafrika. Abhandlungen der hamburgischen Kolonial-Instituts, Bd 34. Reihe B: Völkerkunde, Kulturgeschichte und Sprachen, Bd 19. Hamburg: Friederichsen, De Gruyter & Co. Pp. 180, 1916.

Dickens 1994 = *Patrick J. Dickens.* English-Ju|'hoan, Ju|'hoan-English dictionary. Quellen zur Khoisan-Forschung/Research in Khoisan studies, Bd 8. Köln: Rüdiger Köppe Verlag. Pp 371, 1994.

Doke 1925 = *Clement Martyn Doke.* An outline of the phonetics of the language of the Chu: Bushmen of north-west Kalahari. In: Bantu studies and general South African anthropology, 2 (3), pp 129-165, 1925.

Dornan 1917 = *Samuel Dornan.* The Tati Bushmen (Masarwas) and their language. In: Journal of the Royal Anthropological Institute of Great Britain and Ireland, 47, pp 37-112, 1917.

Ehret 1986 = *Christopher Ehret.* Proposals on Khoisan reconstruction. In: African hunter-gatherers (international symposium), pp. 105-130. Ed(s).: Franz Rottland & Rainer Vossen. Sprache und Geschichte in Afrika, special issue 7(2). Hamburg: Helmut Buske Verlag, 1986.

Ehret 2003 = *Christopher Ehret.* Toward Reconstructing Proto-South Khoisan. [In the current volume.]

Gruber 1975 = *Jeffrey S. Gruber.* Plural predicates in #Hoà̃. In: Bushman and Hottentot linguistic studies, pp 1-50. Ed(s): Anthony Traill. Communications from the African Studies Institute, no 2. Johannesburg: University of the Witwatersrand, 1975.

Haacke 1998 = *Wilfrid Haacke, Eliphas Eiseb.* Khoekhoegowab-English, English-Khoekhoegowab glossary. Windhoek: Gamsberg Macmillan Publishers. Pp 278, 1998.

Hastings 2001 = *Rachel Hastings.* Evidence for the genetic unity of Southern Khoesan. In: Cornell working papers in linguistics, 18, pp 225-245, 2001.

Honken 1988 = *Henry Honken.* Phonetic correspondences among Khoisan affricates. In: New perspectives on the study of Khoisan, pp 47-65. Ed(s): Rainer Vossen. Quellen zur Khoisan-Forschung, Bd 7. Hamburg: Helmut Buske Verlag, 1988.

Honken 1998 = Henry Honken. Types of sound correspondence patterns in Khoisan languages. In: Language, identity and conceptualization among the Khoisan, pp 171-193. Ed(s): Mathias Schladt. Quellen zur Khoisan-Forschung/Research in Khoisan studies, Bd 15. Köln: Rüdiger Köppe Verlag, 1998.

Kagaya 1993 = *Ryohei Kagaya.* A classified vocabulary of the Sandawe language. Asian and African lexicon series, vol 26. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa (ILCAA), Tokyo University of Foreign Studies, 1993.

Meinhof 1930 = *Carl Meinhof.* Der Koranadialekt des Hottentottischen. Zeitschrift für Eingeborenen-Sprachen, Beiheft 12. Berlin & Hamburg: Verlag von Dietrich Reimer (Ernst Vohsen); C Boysen. Pp 152, 1930.

Nakagawa 1996 = *Hirosi Nakagawa.* A first report on the click accompaniments of |Gui. In: Journal of the International Phonetic Association, 26 (1), pp 41-54, 1996.

Sands 1998 = *Bonny Sands.* Eastern and Southern African Khoisan. Evaluating claims in distant linguistic relationships. Ed(s): Rainer Vossen, Quellen zur Khoisan-Forschung/Research in Khoisan studies, Bd 14, pp. 256, 1998.

Snyman 1997 = *Jan W. Snyman.* A preliminary classification of the !Xũũ and Žhu|'hõasi dialects. In: Namibian languages: reports and papers, pp. 21-106. Ed(s): Wilfrid [Heinrich] [Gerhard] Haacke & Edward D[erek] Elderkin. Namibian African studies, vol 4. Cologne: Rüdiger Köppe Verlag for the University of Namibia, 1997.

Tanaka 1978 = *Jiro Tanaka.* A San vocabulary of the central Kalahari: G||ana and G|wi dialects. African languages and ethnography, vol 7. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa (ILCAA), Tokyo University of Foreign Studies. Pp xxvii, 158, 1978.

Traill 1973 = *Anthony Traill.* "N4 or S7": another Bushman language. In: African studies, 32 (1), pp 25-32. Johannesburg, 1973.

Traill 1986 = *Anthony Traill.* Do the Khoi have a place in the San? New data on Khoisan linguistic relationships. In: African hunter-gatherers (international symposium), pp 407-430. Ed(s): Franz Rottland & Rainer Vossen. Sprache und Geschichte in Afrika, special issue, 7(1). Hamburg: Helmut Buske Verlag, 1986.

Traill 1994 = *Anthony Traill.* A !Xóõ dictionary. Quellen zur Khoisan-Forschung/Research in Khoisan studies, Bd 9. Köln: Rüdiger Köppe Verlag. Pp 292, 1994.

Vossen 1997 = *Rainer Vossen.* Die Khoe-Sprachen. Ein Beitrag zur Erforschung der Sprachgeschichte Afrikas. Quellen zur Khoisan-Forschung/Research in Khoisan studies. Bd. 12. Köln: Rüdiger Köppe Verlag. Pp. 536, 1997.

Vossen 1992 = *Rainer Vossen.* Q in Khoe: borrowing, substrate or innovation?. In: African linguistic contributions (Festschrift Westphal), pp. 363-388. Edited by: D. F. Gowlett. Helmut Buske, Hamburg, 1992.

Westphal 1965 = *Ernst Oswald Johannes Westphal.* Linguistic research in SWA and Angola. In: Die ethnischen Gruppen Südwestafrikas, pp 125-144. Wissenschaftliche Forschung in Südwestafrika, Bd 3. Windhoek: Südwestafrikanische Wissenschaftliche Gesellschaft, 1965.

Westphal 1980 = *Ernst Oswald Johannes Westphal.* The age of 'Bushman' languages in southern Africa. In: Bushman and Hottentot linguistic studies 1979, pp 59-79. Ed(s): Jan W[inston] Snyman. Pretoria: University of South Africa (UNISA), 1980.

Evolution of Human Languages Project
Santa Fe Institute

1399 Hyde Park Road
Santa Fe, New Mexico 87501
U.S.A.
http://www.santafe.edu/

There are currently about 6,000 languages spoken on our planet, some of them used by millions and some only by a few dozen people. A natural task of the researchers is to provide a detailed classification of those languages, organizing them into a genealogical tree similar to the accepted classification of biological species. Since all representatives of the *Homo sapiens* species are presumably of common origin, it would be natural to suppose that all human languages also go back to some common source, although verifying this is a goal yet to be achieved. Most existing classifications, however, do not go beyond some 300 to 400 language families that are relatively easily observable. This restriction has its natural reasons: languages must have been spoken and constantly changed for at least 40,000 years (and quite probably more), while any two languages separated from a common source inevitably lose almost all superficially common features after some 6,000-7,000 years.

Nevertheless, despite widespread skepticism and reluctance, there are a number of scholars who believe that these obstacles are not insurmountable. Research has been going on during the past several decades that appears to indicate that larger genetic sub-groupings are possible, and, indeed, quite plausible. It can be shown that most of the world's language families can be classified as belonging to about a dozen large groupings, or macro-families. Two sorts of evidence can be used for this purpose:

1) Even a superficial analysis of the vocabulary of a large number of linguistic families reveals numerous lexical similarities extending far over the borders of smaller genetic units. They are frequently restricted to individual macro-families (such as Eurasiatic, Afroasiatic etc.), but a significant number of such matches is already found between macro-families themselves, pointing to the probability of common origin.
2) Classical historical linguistics has developed a very powerful tool: the comparative method, which allows us to reconstruct unattested language stages, so-called "proto-languages." It turns out that whereas modern languages may be very significantly changed, proto-languages in various cases tend to be much more similar to each other. This is the case, for example, with Indo-European, Uralic and Altaic: modern English, Finnish and Turkish may have almost nothing in common, but their respective ancestors – Proto-Indo-European, Proto-Uralic and Proto-Altaic – appear to have many more common traits and common vocabulary. This means that it is possible to extend the time perspective and reconstruct even earlier stages of human language. Much of this research has already been conducted.

The amount of information that has to be processed in order to achieve a deep linguistic taxonomy is enormous, keeping in mind that one has to process thousands of languages and hundreds of linguistic families. Modern computer technology, however, provides some solutions for these problems. The first step that needs to be taken is a compilation of computer databases containing established matches between related languages (etymologies). The primary goal of the EHL research is, therefore, collecting and compiling such databases, and making them easily available: in today's world it means making them available online. A large set of computer databases is already available, and many of them are already online. The databases provided by the EHL participants and freely browsable on the Web include: Altaic, Dravidian, North Caucasian, Yeniseian, Sino-Tibetan, Indo-European, Austroasiatic, Chukchee-Kamchatkan, and Semitic. For many other language families the databases are in the stage of preparation.

Etymological databases for several macro-families are also being compiled, and several of them – Australian, Eurasiatic (Nostratic) and Afroasiatic – are already near completion. Once an etymological database becomes available, it can be used to significantly simplify the task of searching lexical cognates and building up higher-level databases. Etymological databases can also be used (and are used) for a statistical evaluation of taxonomic correlations. The number of etymological matches between languages is a good measure of the distance between them, and can also be employed for evaluating the time depth of any linguistic family. In fact, so-called lexicostatistics is the only available tool for absolute linguistic dating, and its theoretical rationale and practical employment are among the central tasks of EHL.

While the project is concentrated around building up a hierarchical system of etymological databases, reflecting the hierarchical taxonomy of the linguistic genealogical tree, it is also concerned with collecting and putting online primary language wordlists, as well as existing etymological sources. The ideal etymological database system should be able to provide an etymology for any word of any modern or ancient language, tracing its origin as far as possible. The participants of the project have provided source wordlists for poorly explored language families like Indo-Pacific and Australian, where most of the comparative work is yet to be done. They have also scanned, recognized and converted to database format some of the major existing etymological dictionaries, such as Pokorny's Indo-European etymological dictionary.

The ultimate goal of the system of databases as described above is to arrive at a stage where an absolute majority of the world's languages can be reduced to a minimum number of huge macro-families, which, in turn, can be traced back to a Proto-World stage, in case the databases provide enough evidence for the hypothesis of monogenesis. With the database system completed, and the basics of the Proto-World structure established, we can hope to come in possession of a vital tool for helping us understand the nature of language origin in itself.

Sergei A. Starostin
Russian State University for the Humanities,
Santa Fe Institute

Linguistic Databases and Linguistic Taxonomy Workshop

Santa Fe Institute

January 6-10, 2003

The Conference began on January 6 with a welcome from the organizers, Murray Gell-Mann and Sergei A. Starostin. Nobel Laureate Murray Gell-Mann is a distinguished fellow at the Santa Fe Institute, and Sergei A. Starostin, a professor at Russian State University of the Humanities in Moscow, has been residing for periods of time at Santa Fe as part of the Evolution of Human Language Project (EHL).

The general plan of the conference consisted of about six presentations each day. Each presenter was allotted an hour for the presentation and discussion. The presentations on Monday through Thursday were as follows:

Jim Mason (Director of the Rosetta Project, San Francisco, CA) updated the conference on the progress of the Rosetta Project. See http://www.rosettaproject.org/live

Merritt Ruhlen (Stanford University and SFI) reported on "The Current State of Linguistic Taxonomy."

Paul Whitehouse, working in London for SFI, spoke on "Inclusion Versus Exclusion: The Problem of Negative Evidence."

Alexander Lubotsky (Leiden University) reported on the progress of the Leiden Indo-European Etymological Dictionary Project.
See http://iiasnt.leidenuniv.nl/ied/index.html

Sergei A. Starostin (Russian State University, SFI) reported on the extensive language database being compiled for the EHL. A compact disk containing the database was distributed to conference participants. See http://starling.rinet.ru/index2.htm

Martine Robbeets (Leiden University) presented a report on her doctoral thesis, a thorough evaluation of the evidence for the hypothesis that Japanese is an Altaic language.

Anna Dybo (Russian Academy of Science, Institute of Linguistics) reported on the historical contacts between the Ainu language of Japan and various Altaic languages.

Aharon Dolgopolsky (University of Haifa) presented the paper "Proto-Nostratic: a synthetic or analytic language."

Harold C. Fleming (ASLIP and Boston University) reported on "The Grand Strategy" in the search for Mother Tongue.

William Baxter (University of Michigan) reported on "New techniques for reconstructing the history of Chinese 'dialects'."

John D. Bengtson (ASLIP and SFI) presented a paper on "Basque comparative phonology."

Vitaly Shevoroshkin (University of Michigan) reported on "Salishan and North Caucasian."

Timothy Usher (Rosetta Project and SFI) reported on recent comparative work testing the validity of Greenberg's Indo-Pacific hypothesis, and on the huge Indo-Pacific database being compiled.

Ilya Peiros (Max Planck Institut and SFI) reported on the Intercontinental Dictionary Series and historical linguistics of Southeast Asia.
See http://ves101.uni-muenster.de/IDS

Dmitry Leshchiner (SFI) presented the paper "Hokan Comparative Studies – Status and Prospects in Larger Amerind Context."

Luca Cavalli-Sforza (Stanford University) reported on the latest information on the evolution of modern humans, according to population genetics.

Václav Blažek (Masaryk University and Brno Institute of Linguistics) presented a report on Afro-Asiatic glottochronology.

Christopher Ehret (University of California/ Los Angeles) presented his results on the reconstruction of the Proto-South-Khoisan and Proto-Khoisan proto-languages.

George (Georgiy) Starostin (Center of Comparative Linguistics, Moscow, and SFI) presented a paper "Towards a Reconstruction of Proto South African Khoisan."

Alexander Militarev (Moscow Jewish University) reported on "Dating Proto-Afro-Asiatic."

Some others who did not have specific presentations but participated in organized and spontaneous discussions, or otherwise participated behind the scenes were Bernard Comrie (Max Planck Institute), William S.Y. Wang (City University of Hong Kong), Vittorio Loreto (La Sapienza University, Rome), Natalie Operstein (Los Angeles), Lisa Diamond, and Kurt Bollacker.

This conference was a long-delayed sequel to the Symposium on Language and Prehistory, held at Ann Arbor in 1988. At least 11 of the participants at the Ann Arbor conference were re-united for the present conference. For some of the Russians in 1988 it was their first journey to the United States and a first taste of the freedom brought about by the fall of the Iron Curtain. Now, almost 15 years later, we can report on significant progress in the study of language in prehistory. Much of this progress was made possible by the contacts between Western and Soviet-Bloc scholars initiated by Hal Fleming and Aharon Dolgopolsky in the early 1980's. They (along with Gell-Mann, Starostin, Ruhlen, Wang, Comrie) deserve much of the credit and thanks for the state of the art in paleo-linguistics.

John D. Bengtson
Vice President
Association for the Study of Language in Prehistory